Universiteit Antwerpen

Faculteit Wetenschappen

Departement Fysica

Data-driven methods for the analysis of time-resolved mental chronometry fMRI data sets

Approaches and techniques based on the Fuzzy Clustering Method and spatial Independent Component Analysis

Data-gedreven methodes voor de analyse van tijds-geresolveerde mentale chronometrie fMRI data

Aanpak en technieken gebaseerd op de Vage Clustering Methode en spatiale Onafhankelijke Componenten Analyse

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de Wetenschappen
aan de Universiteit Antwerpen te verdedigen door

Alain Smolders

Promotor
Prof. Dr. Jan Sijbers
Copromotors
Prof. Dr. Paul Scheunders
Prof. Dr. Elia Formisano

Antwerpen, 2007

Data-driven methods for the analysis of time-resolved mental chronometry fMRI data sets

Approaches and techniques based on the Fuzzy Clustering Method and spatial Independent Component Analysis

Alain Smolders

PhD Dissertation

2007

University of Antwerp

Promotor:
Prof. Dr. Jan Sijbers
Co-promotors
Prof. Dr. Paul Schuenders
Prof. Dr. Elia Formisano

— *To ...* —

— ... —

Acknowledgements

Contents

Summary

This dissertation deals with data-driven methods for the analysis of complex Blood Oxygen Level Dependent (BOLD) functional Magnetic Resonance Imaging (fMRI) data sets. It explores and identifies the spatial layout and sequence of brain activations associated with cognitive tasks. The relationship between the measured haemodynamic effect and the underlying neuronal activity however, is complex and poorly understood. Moreover, the resulting signal change is small and exhibits a nonlinear component. Finally, confounding factors and the disturbance of the weak signal by several noise sources hamper the interpretation of the results. These constraints become particularly important when analysing complex, cognitive data sets. However, several approaches exist to deal with these constraints, such as a well-considered choice of experimental design, acquisition technique, and analysis method. In this dissertation, the focus lies on the analysis method.

Originally, hypothesis-driven analysis techniques were applied. They specify a priori a spatially-invariant model of the haemodynamic response (HR). Its goodness-of-fit is tested at each voxel by statistical methods. However, the assumption of spatial invariance of the HR is suboptimal for the analysis of complex tasks. The latter involve the activation of extended networks of brain regions with widely different HRs, exhibiting a substantial degree of trial-by-trial variability. Moreover, hypothesis-driven methods typically ignore interactions between voxels. Therefore, an alternative, data-driven approach was introduced. This shift led to the application of the Fuzzy Clustering Method (FCM) and spatial Independent Component Analysis (sICA). In both methods the data are decomposed into a set of spatio-temporal modes, without strong a prior assumptions about the temporal profile of the effects of interest. These methods however have a different view on the data. Spatial ICA adopts a spatial view on the data, i.e. the data are considered as a sequence of volumes, whereas FCM adopts a temporal view, i.e. the

data are considered as a spatial distribution of time courses. FCM divides the time courses into a few types of activation, expressing the underlying neural activity. Each group or cluster is characterised by its spatial map and corresponding time course, the cluster centre. FCM expresses the similarity of a voxel time course to a cluster centre by a so-called membership to that cluster. Spatial ICA is an information theoretic approach, searching for statistical independence between the components. It models the data as a linear combination of independent components (ICs), expressing the underlying neural activity.
Both methods exhibit specific shortcomings for the analysis of complex, cognitive data sets. A well-considered selection of method-specific parameters and the correct validation of the results is difficult to accomplish. Moreover, limitations on the inherent detection accuracy of the methods may hamper the correct dissection of the processing stages of complex, cognitive tasks.

In this dissertation, two approaches based on FCM and sICA are introduced to overcome these shortcomings. Firstly, we deduce guidelines for the choice between the methods and for the setting of their parameters in the context of a complex, cognitive task. Secondly, we present an enhanced FCM algorithm that incorporates spatial information in the detection process. This approach takes into account that task-induced neuronal activations and BOLD responses are expected to produce similar signal changes in spatially contiguous regions, extending over several millimetres.
The evaluation and comparison of these approaches and algorithms is performed in terms of the detection accuracy and consistency, the robustness, the ease in determination of parameter settings, and the validability and interpretability of the results. In order to assess these criteria, appropriate data sets are presented. Therefore, a newly devised complex visuospatial mental imagery experiment is introduced, yielding time-resolved mental chronometry functional Magnetic Resonance Imaging (fMRI) data sets.

In general, two approaches to accomplish this assessment exist. In order to assess the aforementioned criteria *quantitatively*, simulations and tests on a *synthetic* data set are indispensable. The assessment on *real* data sets on the other hand, can only be performed *qualitatively*. However, findings resulting from earlier research, applying alternative analysis techniques to the same (or similar) data sets, offer a reference framework. Both approaches are applied in this dissertation, in two separate studies.

A first study provides empirical guidelines for the choice between both methods, i.e. FCM or sICA, as well as for their practical use. Both algorithms are evaluated and compared on real fMRI data sets, acquired in the context of the visuospatial mental imagery experiment. Validability and interpretability of physiologically meaningful components, as well as robustness of the method are assessed qualitatively. Detection accuracy and consistency of the resulting time courses and spatial maps within and between-subjects are assessed and compared quantita-

tively, using the temporal and spatial correlation, respectively. In this study, the influence on the latter assessment criteria of both the method-specific parameter settings and preliminarily averaging over trials is determined as well.

A second study incorporates neighbourhood information in the 'conventional' FCM algorithm. It elucidates under which conditions and for which parameter settings 'spatio-temporal' FCM outperforms conventional FCM in terms of detection accuracy. Therefore, realistic simulated data sets are constructed, holding different types of activation synthetically embedded in a realistic noise background. The latter data sets are generated for a range of CNR values normally encountered in fMRI analysis. A quantitative comparison between activation regions as well as the corresponding time courses obtained by both methods is performed. The latter is accomplished using the Receiver Operating Characteristics (ROC) methodology and the temporal correlation, respectively.
Finally, the findings obtained in the context of the simulated data set are applied to the analysis of the visuospatial mental imagery data sets with both methods. This approach allows to determine and interpret the differences between corresponding topological structures obtained by both methods.

Outline

This thesis is divided in two parts. Part I elaborately reviews the requisite background on Blood Oxygen Level Dependent (BOLD) functional Magnetic Resonance (fMRI) measurement of cortical activity (chapter 1) and analysis techniques for fMRI data sets (chapter 2). Part II provides the main contributions, i.e. the proposed approaches and techniques, the visuospatial mental imagery experiment, yielding appropriate data sets (chapter 3), and the applied studies (chapter 4 and 5).

Chapter 1 expounds on the relationship between the measured haemodynamic effect and the underlying neuronal activity. It focuses on the subsequent difficulties for the analysis of complex, cognitive tasks. It considers the confounding factors and the disturbance of the weak signal by several noise sources. Furthermore, this chapter elaborates on the impact of the whole of these effects on the specificity of the BOLD signal. Finally, approaches to improve the specificity by appropriate paradigms, experimental design and acquisition techniques are proposed.

Chapter 2 covers the techniques to analyse fMRI data sets. It explains the shift from a hypothesis-driven approach to a data-driven approach. It covers into detail two data-driven techniques currently favoured for the analysis of complex data sets: the Fuzzy Clustering Method (FCM) and spatial Independent Component Analysis (sICA). It clarifies their principles, emphasises their strengths and shortcomings and presents a literature review of the major applications in a historical perspective. Finally, the Receiver Operating Characteristics (ROC) technique to assess the inherent accuracy of a detection procedure is explained and applied in the fMRI context.

Chapter 3 deals with FCM and sICA based approaches for the analysis of time-

resolved event-related mental chronometry data sets. It elucidates the dual view on the data, i.e. temporal or spatial. It elaborates into detail on several algorithms implementing FCM and sICA and introduces new approaches and algorithms based on both techniques. Chapter 3 furthermore presents an appropriate data set allowing to evaluate these approaches and techniques when dissecting the distinct processing stages of a complex cognitive task. Therefore, a visuospatial mental imagery experiment is introduced, yielding a multi-subject, multi-run, time-resolved event-related mental chronometry fMRI data set.

Chapter 4 evaluates and compares FCM and sICA for the dissection of cognitive stages of the visuospatial mental imagery experiment. It provides empirical guidelines as to choose between FCM and sICA, as well as for their use. The comparison holds a qualitative evaluation of the time courses and spatial maps obtained by both methods. Moreover, it quantitatively evaluates the within- and between subject-consistency as well as the spatial and temporal correspondence between the decompositions obtained by both methods.

Chapter 5 compares 'conventional' FCM to 'spatio-temporal' FCM when analysing fMRI time series. It elucidates the concept of 'spatio-temporal' FCM and compares both methods on realistic simulated data sets, generated for a range of CNR values normally encountered in fMRI analysis. It investigates under which conditions and for which parameter settings spatio-temporal FCM outperforms conventional FCM in terms of detection accuracy. Finally, both methods are applied to the visuospatial mental imagery data sets, applying the findings obtained in the context of the simulated data sets. The differences between corresponding topological structures obtained by both methods are determined and interpreted.

Terminology

Arterial spin labelling (ASL). ASL is an imaging technique based on cerebral blood perfusion and uses RF pulses to 'label' flowing blood (i.e. the longitudinal magnetisation is flipped) without the need for an exogenous contrast agent. Although ASL techniques have high spatial and temporal precision, the methods are not in widespread use for functional imaging because they provide lower (half or even less) signal-to-noise than BOLD-based fMRI.

Blood composition (human). Human blood is composed of plasma (55%) and cells (45%). Blood plasma contains proteins but mainly consists of water, composed of oxygen and hydrogen nuclei. Hydrogen nuclei (protons) possess an intrinsic magnetic moment and are therefore experiencing the NMR effect. The cell component is dominated by red blood cells, containing haemoglobin.

Cerebral blood flow (CBF) is defined as the volume of blood passing per unit time through the capillary bed in a given mass of tissue and is commonly expressed in 'millilitres of blood per 100 g of tissue per minute'. Because the density of the brain tissue is close to 1 g/ml, CBF is also expressed in units of inverse time. A typical average value for CBF in the human brain is 60 ml/100 (g min) or 0.01 s^{-1}.

Cerebral blood volume (CBV) is defined as the volume of blood present in a given quantity of brain tissue and is dimensionless. A typical value for CBV in the human brain is 4 %. Estimates of the relative sizes of CBV are: 5 % for the arterial volume, with the rest divided about equally between capillaries and veins.

Cerebral cortex. See **Grey matter**.

Cerebral metabolic rate of glucose (CMRGlc). See Metabolism

Cerebral metabolic rate of oxygen ($\mathbf{CMRO_2}$). See Metabolism

Cerebrospinal fluid (CSF). See **Grey matter**.

Diffusion. The diffusion effect, also known as Brownian motion, describes that a group of molecules starting at the same position, will spread over time yielding a displacement with zero mean and standard deviation growing as the square root of time. The standard deviation of a water molecule in the blood is typically $20\,\mu$m for a time period of 200 ms. Capillaries typically have a diameter of less than $10\,\mu$m and a blood speed of 1 mm/s, small veins typically have a diameter of at least $60\,\mu$m and a blood speed of at least 5 mm/s.

Computed Tomography (CT) is an imaging technique measuring the local X-ray attenuation coefficient. Digital geometry processing is used to generate a three-dimensional image of the internals of the subject from a large series of two-dimensional slices taken around a single axis of rotation.

Electroencephalography (EEG) is the neurophysiologic measurement of the electrical activity of the brain by recording from electrodes placed on the scalp or, in special cases, subdurally or in the cerebral cortex. The resulting traces are known as an electroencephalogram (EEG) and represent an electrical signal from a large number of neurons.

Gradient echo (GE). See **Pulse sequence**.

Grey matter The brain consists of grey matter (40%) and white matter (60%) contained within the skull. It is surrounded by Cerebrospinal fluid (CSF), cushioning the brain, and is connected to the spinal cord, which carries nerve messages for sensation and movement between the brain and the body. Generally, white matter can be understood as the parts of the brain responsible for information transmission, whereas grey matter is responsible for information processing. The cerebral cortex is the extensive outer layer (typically 1-4mm thick) of grey matter, is rich in neurons and therefore largely responsible for higher brain functions.

Haemodynamics refers to the branch of physiology that studies the circulation of blood through the cardiovascular system and the pressures in the heart.

Haemoglobin is a protein in the red blood cells for the transportation of oxygen. Haemoglobin consists of 4 polypeptide chains, which are identified in the most common type of adult haemoglobin as alfa and beta chains. When bound to oxygen, it is called oxyhaemoglobin (O_2HbFe^{2+}). When oxygen is released, it is called deoxyhaemoglobin ($HbFe^{2+}$).

Magnetoencephalography (MEG) is the neurophysiologic measurement of the

magnetic fields produced by the electrical activity of the brain. The resulting traces are known as an magnetoencephalogram (MEG) and represent a magnetic signal from a large number of neurons.

Metabolism is the sum of chemical and physiological processes in living organisms by which substances are produced or transformed to produce energy. In the brain, oxygen and glucose are consumed as to supply the energy for the basic processes of cellular work. The corresponding rates are called 'Cerebral Metabolic Rate of Oxygen' ($CMRO_2$) and 'Cerebral Metabolic Rate of Glucose' (CMRGlc), respectively.

Neural system (nervous system). The system that coordinates the activity of the muscles, monitors the organs, constructs and processes input from the senses, and initiates actions. It consists of the Central Nervous system (CNS) and the Peripheral Nervous System (PNS). The CNS is located within the skull and spine, whereas the PNS is located outside the skull and spine.

Neurovascular refers to blood vessels of the neural system.

Perfusion is a physiological term that refers to the process of nutritive delivery of glucose and oxygen of arterial blood to a capillary bed in the biological tissue. In medical imaging perfusion implies looking at phenomena such as regional cerebral blood flow and volume, mean transit time, time to peak, and capillary permeability.

Physiology is the branch of biological sciences dealing with the functioning of living organisms or any of its parts (organs, tissues, cells).

Positron emission tomography (PET) is a nuclear imaging technique. A short-lived radioactive tracer isotope, which is incorporated into a metabolically active molecule, is injected in the blood circulation of the subject. The metabolically active molecule becomes concentrated in tissues of interest (e.g. the brain). The isotope decays by emitting a positron, an effect used to produce a three-dimensional image of functional processes.

Pulse sequence. A pulse sequence is a set of RF pulses applied to a sample to produce a specific form of MR signal. Conventional fMRI imaging techniques consist of two broad classes: spin echo (SE) and gradient echo (GE). An SE technique includes a 180° RF refocusing pulse that corrects for signal loss due to magnetic field inhomogeneities. A GE technique has no refocusing pulse, yielding a much shorter repetition time at the expense of a larger attenuation of the acquired signal, however. Echo Planar Imaging (EPI) is a fast imaging technique.

Spin echo (SE). See **Pulse sequence**.

Transcranial magnetic stimulation (TMS) is a noninvasive method to excite neurons by applying weak electric currents induced in the tissue by rapidly changing magnetic fields. In repetitive TMS (rTMS), stimuli are repetitively applied.

White matter. See **Grey matter**.

Notations

Let $\boldsymbol{X}$ be a $T \times N$ matrix of observations.

$\boldsymbol{x}_t$ is the t-th row vector of observations, with dimensions $1 \times N$.

${\boldsymbol{x}_n}^T$ is the n-th column vector of observations, with dimensions $T \times 1$.

> A column vector is denoted with the transpose operator $(.)^T$. Whenever an elaboration applies a deviating notation, such as with SVD, the latter will be mentioned explicitly.

$x_{t,n}$ is the observed matrix element on row t and column n.

$\widetilde{\boldsymbol{x}}_t$ is a vector of random variables.

$\widetilde{\boldsymbol{X}}$ is a matrix of random variables.

$\bar{\boldsymbol{X}}$ is the mean over the rows of matrix $\boldsymbol{X}$.

$\bar{x}$ is the mean of vector $\boldsymbol{x}$.

$\boldsymbol{X}_{\mathrm{w}}$ is the whitened transform of matrix $\boldsymbol{X}$.

$C = \mathrm{E}\left\{\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T\right\}$ is the covariance matrix of random matrix $\widetilde{\boldsymbol{X}}$.

> $\mathrm{E}\{.\}$ is the expectation-operator.

$\widehat{\boldsymbol{C}} = \boldsymbol{X}\boldsymbol{X}^T$ is the estimate of $\boldsymbol{C}$ based on the observed matrix $\boldsymbol{X}$.

Part I

DATA-DRIVEN ANALYSIS OF BOLD FMRI DATA SETS

Chapter 1

BOLD fMRI measurement of cortical activity

1.1 Introduction

This chapter elaborates on the relationship between the measured haemodynamic effect and its underlying neuronal activity. It focuses on the complexity of this relationship and the fact that it is poorly understood. Furthermore, the confounding factors and the disturbances of the weak signal by several noise sources are elucidated. The impact of these effects on the specificity of the BOLD signal is considered as well as approaches to improve the specificity.

1.2 Functional imaging techniques

1.2.1 Historical overview

First approaches

The goal of understanding the functional organisation of the human brain has motivated neuroscientists for well over 100 years, but the experimental tools to measure and map brain activity have been slow to develop. Fluctuating electric and magnetic fields measured at the scalp with electroencephalography (EEG) or magnetoencephalography (MEG), respectively, provide information on electrical events within the brain. From these data, the location of a few sources of activity is estimated, but the information is not sufficient to produce a detailed map of the pattern of activation.

The idea that cerebral blood flow (CBF) could reflect neuronal activity in the concerning brain regions offered a promising alternative [1]. Neuronal activity could now be determined by measuring the precise location of the metabolic activity that followed. In 1948, Kety and Schmidt described the nitrous oxide technique for measuring CBF [2]. It was the first technique capable of producing quantitative measurements of global brain blood flow in humans, providing a new window on the physiological functioning of the brain.
Subsequent techniques with radioactive tracers, like the radioactive xenon technique [3], led to positron emission tomography (PET) [4, 5], which is able to measure CBF locally with improved spatial resolution. Patterns of activation in the working human brain could now be mapped.

Nuclear Magnetic Resonance

In 1946, two teams led by Purcell and Bloch independently and simultaneously discovered the Nuclear Magnetic Resonance (NMR) effect [6, 7]. Both research groups detected that certain nuclei (including hydrogen) possess an intrinsic magnetic moment, and when placed in an external magnetic field, they orient themselves along this external magnetic field, performing a precession with a frequency proportional to the field strength. This finding had far-reaching applications in many domains, such as Magnetic Resonance Imaging (MRI), which was developed in the 1970s as a new tool to reveal details of human anatomy. When a person is placed in an MRI scanner, the aforementioned orientation of the intrinsic magnetic moment of the hydrogen nuclei is disturbed by exciting these nuclei with a burst of electromagnetic energy in the form of radiofrequent (RF) pulses. Afterwards, the nuclei realign themselves by transmitting an RF signal that is detected in a receiver coil placed around the patient's head. The resulting signal reflects the tissue-dependent magnetization of water molecules in space and can thus be used to produce an image. In 1973, Lauterbur was the first to introduce an imaging technique based on MR [8]. MRI quickly became an indispensable tool in diagnostic radiology with a large flexibility: contrast between one tissue and another in an image is varied simply by varying the acquisition parameters. Although very interesting, a detailed description of the basic physics of MRI and the way the MR signal is manipulated experimentally in the formation of an image is considered beyond the scope of this dissertation and is therefore omitted. An excellent and complete description of the principles of MR and the MR imaging techniques is found in the work of Buxton [9]. In this dissertation, we will frequently refer to several issues covered by these imaging techniques, such as the applied pulse sequences, the tissue- and pulse sequence dependent parameters, and the corresponding contrast characteristics.

Functional Magnetic Resonance Imaging

In the late 1980s, MR led to a revolutionary technique for research on the basic functions of the human brain: functional magnetic resonance imaging (fMRI). It was found that the MR signal's sensitivity to local changes in perfusion allowed the detection of local brain activity simply by evaluating the corresponding MR signal change. Originally, the MR signal was made sensitive to perfusion by intravascular contrast agents, such as gadolinium- diethylene-triamine-pentaacetic acid (Gd-DTPA) [10]. In the early 1990s, however, Ogawa et al. discovered a natural physiological contrast mechanism. They found that local changes in the oxygenation of the blood affected the MR signal and called this phenomenon the 'Blood Oxygenation Level Dependent' (BOLD) effect [11]. This finding resulted in the first experiment measuring activity in the human brain, by Kwong et al. [12].

1.2.2 Strength and shortcomings of fMRI

Strength

The flexibility of fMRI makes it a powerful tool to identify anatomic sources of the electrophysiological events together with their temporal behaviour. Although comparatively new, fMRI is now a primary technique for basic studies of the organisation of the working human brain, because of several advantages over PET:

(1) fMRI has in principle higher spatial and temporal resolution.

(2) BOLD fMRI procedures are completely non-invasive and can be repeated in a single subject without concern for exposure to ionizing radiation.

(3) Since anatomic and functional images can be acquired during the same imaging session, functional maps can be superimposed directly to anatomic images without any misregistration.

Figure 1.1 shows a typical fMRI image in which a coloured activation map is overlaid on a high resolution anatomical image of a transversal slice, recorded during the same session. In the (T_1-weighted, see [9]) anatomical image, grey matter is dark grey, white matter light grey, cerebrospinal fluid (CSF) is black. The activation map indicates regions of auditory activation and a lighter shade of a colour refers to a more significant result found by the applied method (see section 4.2.5).

Shortcomings

As a disadvantage of fMRI, neuronal activity is not measured directly, but instead by means of the haemodynamic effect associated with activation through a neurovascular coupling composed of a cascade of complex and poorly understood mechanisms [14, 15]. This limits the temporal resolution of fMRI to be orders of

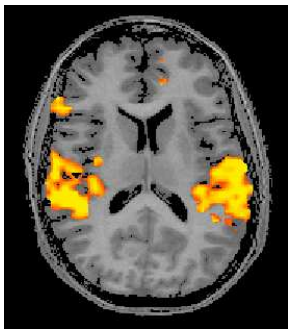

Fig. 1.1: *Typical fMRI image in which a coloured activation map is overlaid on a high resolution anatomical image of a transversal slice, recorded during the same session [13].*

magnitude poorer than that of EEG or MEG, and the resolution in space faces a natural blurring and misregistration of the spatial sources of the BOLD signal. The latter differs slightly from the neuronal source registration [16], since haemodynamic effects can be somewhat dislocated to the neuronal events.

Moreover, fMRI's flexibility makes it a tool difficult to grasp. A computed tomography (CT) X-ray image, for example, is relatively easy to understand: the intensity at each point in the image is proportional to the corresponding local X-ray attenuation coefficient. But in MRI, the intensity at each point in the image depends not only on the local proton density but also on intrinsic tissue properties, blood oxygenation, blood flow, the heterogeneous structure of the tissue, and the local diffusion characteristics of water including anisotropic directional properties. The extent to which each of these physical characteristics affects the MR signal depends on how exactly the image is acquired.

1.3 Functional relationship between BOLD response and underlying neuronal activity

1.3.1 Pooled neuronal activity

While one cannot non-invasively examine the response of single neurons in the human cortex, a number of investigators examining the visual cortex [17–20] have argued that behaviourally or perceptually significant activity occurs not at the level of single cells, but in large pools of neurons. Consequently, fMRI, measuring activity at the voxel level (typically 30 mm^3), applies to the study of human behaviour. However, there are several underlying assumptions implicit in any assertion that fMRI reflects behaviourally relevant neuronal responses.

- fMRI response from a selected region of cortex has to be proportional to the

spatial average of individual neurones' metabolic activity integrated over some time [14, 21].

- Physiological judgements have to be governed by pooled neuronal activity and not by one or a small number of stimulus-specific neurones [22, 23].

- A certain number of neurones must spike above a threshold mean rate in order for behaviour to occur [24].

1.3.2 Physiological changes during brain activation

Several physiological changes occur due to neuronal activity, such as a change in energy metabolism, cerebral blood flow (CBF), and cerebral blood volume (CBV). We expound on their relation and their use as to estimate relative changes in the oxygen consumption of brain tissue, accompanying neuronal activity.

Cerebral blood flow and energy metabolism

Under normal conditions, the brain derives almost all of its energy from the oxidation of glucose. Therefore, it needs a nearly constant supply of glucose and oxygen, delivered by the blood supply through a rich network of vessels. Although the brain accounts for only about 2 % of the total body mass, it consumes 20 % of the body's glucose and oxygen and receives 20 % of its blood supply. Energy metabolism and CBF are tightly linked to local neuronal activity, which implies that maps of local glucose consumption, local oxygen consumption, or local blood flow each provide information on neuronal activity. A complete description of processes involved in neuronal activity from the perspective of thermodynamics is found in Nicholls et al. [25].

Cerebral blood volume

Cerebral blood volume is a distinctly different physiological quantity compared to CBF. Both are related however, because the change in CBF is accomplished by a change in the cerebrovascular resistance, which is concentrated in the arterioles. A dilation of the arterioles leads to a reduction in the resistance of the arterioles and a reduced pressure drop across the arteriolar segment. This will increase the pressure in all later segments of the vascular tree and dilates the veins (and possibly the capillaries as well). The result is that changes in the blood volume are not likely to be evenly distributed along the entire vascular tree, but may chiefly be accomplished by the veins. The qualitative relationship between CBF and CBV changes is derived in the work of Grubb et al. [26]:

$$\frac{\mathrm{CBV}}{\mathrm{CBV}_0} = \left(\frac{\mathrm{CBF}}{\mathrm{CBF}_0}\right)^{\alpha} \quad , \tag{1.1}$$

where α is a constant with an approximate empirically derived value of 0.4 and the subscript '0' refers to the baseline steady state of a variable.

1.3.3 Magnetic susceptibility changes accompany metabolic changes

Exogenous contrast agent

Cerebral blood volume was first measured with the use of exogenous contrast agents, such as gadolinium-DTPA (Gd-DTPA) [27, 28]. The presence of any substance in the magnetic field alters that field to some extent and the degree of this effect is called the 'magnetic susceptibility'. Gadolinium-DTPA, which has several unpaired electrons, largely disturbs the local magnetic field, leading to a large magnetic susceptibility effect. This in turn leads to attenuation of the MR signal. Although the agent is confined to the intravascular space, the total MR signal is affected because the microscopic field gradients penetrate into the extravascular space, significantly extending beyond the vessel wall. The resulting signal changes can be quite large (30-50 %) and grow for larger blood volumes.

Following a bolus injection of the agent, the local MR signal in the brain drops transiently as the agent passes through the vasculature. This effect lasts only a brief time (about 10 s) and fast dynamic imaging is thus required to measure it. By integrating over the first passage of the contrast agent, cerebral blood volume is determined. A bolus of contrast agent is injected twice, namely when the subject is at rest and while performing a task and the relative blood volume maps are compared to infer which areas of the brain have been activated. The first fMRI measurement of human brain activation using Gd-DTPA as a contrast agent was performed by Belliveau et al. in a visual stimulus experiment [10].

The major shortcomings of this fMRI technique are the poor temporal resolution and the requirement for an exogenous contrast agent, which limits the functional measurements that can be performed in humans. Therefore, more recently intravascular contrast agents that can remain at stable concentrations in the blood for several hours have been developed [29, 30]. Additional advantages of the latter technique are the potentially high 'signal-to-noise ratio' (SNR) (see section 1.6.1.1) and the fact that the measured signal is directly proportional to CBV. However, as with Gd-DTPA bolus methods, toxicity limits the number and frequency of studies that can be performed on an individual human subject.

Blood oxygenation

With contrast agents, a susceptibility difference between the intravascular and extravascular space is induced by the experimenter. However, there is also a natural physiological mechanism for producing such a susceptibility difference, resulting

from the degree of oxygenation of the blood. Arterial blood arrives fully oxygenated at the brain. Basal neuronal activity consumes oxygen and thus increases the concentration of deoxyhaemoglobin in the venous blood, thereby altering its magnetic properties. Indeed, oxyhaemoglobin(O_2HbFe^{2+}) is diamagnetic whereas deoxyhaemoglobin ($HbFe^{2+}$) is paramagnetic due to the four unpaired electrons [31]. Magnetic flux is reduced in diamagnetic materials, while paramagnetic materials on the other hand have increased magnetic flux.

Oxyhemoglobin therefore does not alter significantly the regional magnetic field and is thus characterized by a small magnetic susceptibility effect. Deoxyhemoglobin on the other hand largely disturbs the local magnetic field in a region of tissue, leading to a significant magnetic susceptibility effect. An increase in the concentration of deoxyhaemoglobin in the blood thus leads to an increase in the magnetic susceptibility (a proportional effect [32]) which results in a decrease of the acquired MR signal. This phenomenon, by which local changes in the oxygenation of the blood effect the MR signal, is called the 'Blood Oxygenation Level Dependent' (BOLD) effect and was first observed by Ogawa et al. [11].

1.3.4 BOLD effect

Imbalance between CBF and $CMRO_2$

Empirically, the coupling between CBF change and functional activity is reasonably tight, both in location and degree of change. This was already indicated by an early PET study, performed by Fox and Raichle, measuring the blood flow response of the visual cortex to photic stimuli of different frequencies [33]. In a later repetition of this experiment applying BOLD fMRI, Kwong et al. found that the BOLD signal closely followed the CBF change, suggesting that the change in blood oxygenation was altered in parallel with the CBF change [12]. Later, convincing evidence for a close relationship between the CBF and the $CMRO_2$ in the visual cortex was reported [34, 35]. However, Kwong et al. also demonstrated that brain activation led to a local signal increase rather than a decrease, which was surprising because it was already demonstrated that a reduction in blood oxygenation led to a signal decrease [36]. Consequently, their measurement surprisingly suggested that the blood becomes more oxygenated with activation.

Fox and Raichle on the other hand were the first to discover this imbalance between some basic physiological changes following brain activation [5]. They performed a somatosensory experiment and found a local CBF increase of about 30 % in the appropriate area of the brain, but only a 5 % increase of cerebral metabolic rate of oxygen ($CMRO_2$). In a later visual stimulation experiment, this imbalance between CBF and $CMRO_2$ changes was confirmed [37]. In recent years, fMRI has provided ample confirmation that the imbalance of flow and oxygen metabolism changes is indeed a physiological phenomenon [38].

Coupling between CBF, CBV, and $CMRO_2$

Several possible explanations for this large increase in CBF accompanying a smaller increase in $CMRO_2$ and the consequent signal changes have been proposed.

(1) A first hypothesis is that CBF is controlled on a coarse spatial scale, with areas of increased $CMRO_2$ occurring on a finer scale. This hypothesis has been poetically described as 'watering the garden for the sake of one thirsty flower' [39]. If confirmed, this model has important consequences for the ultimate spatial resolution obtainable with functional neuroimaging techniques based on blood flow changes. The limitation on spatial resolution may then be physiological, rather than technological, determined by the minimum volume of tissue over which CBF can be controlled. However, further studies are required to demonstrate that coarse control of CBF can provide a general explanation for the observed imbalance of flow and oxygen metabolism changes.

(2) Another hypothesis for this imbalance is the limited oxygen delivery at rest [40]. A large flow change is required to support a small increase in oxygen metabolism. The argument is based on two assumptions. The first is that oxygen delivery to tissue is 'barrier-limited', so that only a fraction of the oxygen delivered to the capillary bed manages to leave the capillary and become available for metabolism [41, 42]. The second assumption is that a CBF increase is accomplished by increasing the velocity of capillary blood rather than capillary recruitment, thus leading to decreased capillary transit time [43, 44]. This yields that a small CBF *increase* leads to a *reduction of oxygen extraction*. The flow must thus be increased substantially to support a small increase in the oxygen metabolic rate.

(3) More recently, a hypothesis was developed based on dynamics of the BOLD signal change and including the cerebral blood volume (CBV). When a brain region is activated, the oxygen need grows and as a reaction, the cerebral blood flow (CBF) and the cerebral blood volume (CBV) both increase. Several studies indicated CBV changes were slow to follow CBF changes and reported pronounced transient effects of the BOLD signal change [29, 45], as indicated in Figure 1.2 (left). The latter figure shows the dynamics of CBV and CBF signal changes in a rat model, during forepaw stimulation, as measured by Mandeville et al.

    Two similar biomechanical models were proposed to explain this lagging blood volume change: the delayed compliance model [46] and the balloon model [47]. Both models include mechanical properties of the vessels to account for the dynamic changes in blood volume following pressure changes induced by the change in arteriolar resistance. The transient shape of the BOLD signal change follows from these models, keeping in mind that the BOLD effect depends approximately on the total amount of deoxyhaemoglobin within an imaging voxel. This amount depends not only on changes in the oxygen saturation of the blood, i.e. the balance between CBF and $CMRO_2$, but also on the change

in cerebral blood volume (CBV). A CBF increase and a CBV increase have competitive effects on the BOLD signal change. A CBF increase was found to be accompanied closely by a smaller $CMRO_2$ increase [48]. This effect lowers the fraction of deoxygenated blood, thus leading to a positive BOLD signal change. An increase in CBV however increases the amount of blood, and thus increases the amount of deoxyhaemoglobin and so would tend to produce a negative BOLD signal change. Normally, the first effect dominates, leading to a positive BOLD signal. But if the CBV change lags behind the flow change there can be pronounced transient effects, resulting in an initial overshoot of the BOLD signal followed by a reduced plateau and an undershoot at the end of the stimulation. Figure 1.2 (left and right) illustrates the dynamics of the signal changes accounting for the BOLD signal change.

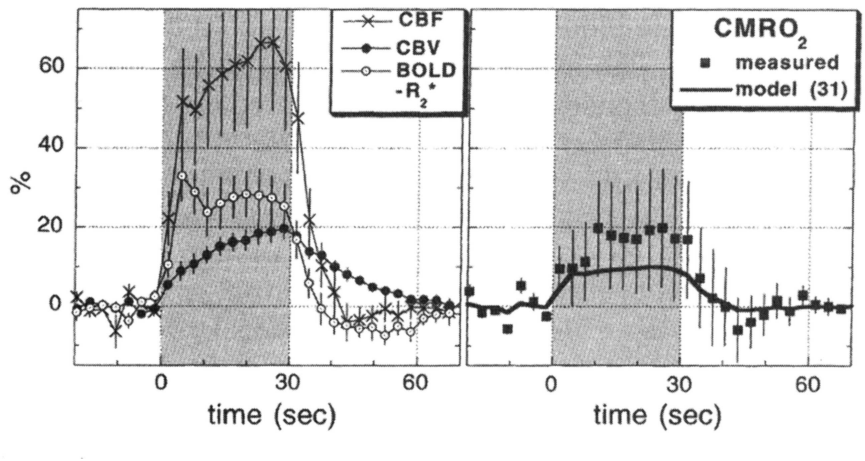

Fig. 1.2: *Haemodynamic changes following stimulation: the CBV lags behind the CBF, leading to a pronounced transient effect in the BOLD signal [48].*

(4) Finally, it is worth mentioning that other possibilities for the explanation of the mismatch between supply and consumption of blood oxygen exist. The vasculature for example could very well deliver a fixed ratio of oxygen and glucose appropriate for an aerobic process. It is suggested that both aerobic and anaerobic processes demand glucose. Because glucose supply appears to match the consumption, the result would be an oxygen surplus [49].

1.4 Characteristics of BOLD signal change

1.4.1 Magnitude of the BOLD signal change

In the resting human brain, arterial blood arrives at the brain fully oxygenated, and around 40 % of the oxygen is extracted in passing through the capillary bed [50]. The venous blood, and to a lesser extent the capillary blood, contains a significant concentration of deoxyhaemoglobin. The susceptibility change due to this amount of deoxyhaemoglobin, observed both in intravascular and extravascular spaces, is about an order of magnitude smaller than that due to a concentrated bolus of Gd-DTPA. This yields a weaker signal attenuation compared to contrast agents, estimated to be about 8 % (field strength 1.5 T) compared to what the signal would be if the blood remained fully oxygenated [51]. When the brain is activated, the venous blood is more oxygenated and the degree of attenuation due to BOLD effect is reduced. The measured signal increase is typically small: a few percent at 1.5 T (At higher fields, e.g. 4 T, the signal changes are much larger, in the range of 5-15 %). For this reason experimenters mostly perform several trials to allow sufficient averaging as to improve the signal-to-noise ratio.

1.4.2 Temporal characteristics of the BOLD response

The haemodynamic response

The haemodynamic response (HR) is the response to a *brief* stimulus. Although neuronal activity can occur very rapidly (on the order of milliseconds) in responses to a sensory event, changes in the HR occur much more slowly (on the order of seconds). As shown in the simple model of Fig. 1.3 a, the HR function is not a simple function of the stimulus pattern, but is temporally blurred in relation to the underlying neuronal activity.

The BOLD response

A simple model for the BOLD response, being the response to a *sustained* stimulus, is presented in Fig. 1.3 b. The latter figure shows a plateau level for the duration of the stimulus, ensuing from two features of the HR: linearity and time-invariance [14]. Boynton et al. assumed that the HR is proportional to the local *average* neuronal activity, averaged over a small region of the brain and averaged over a period of time. Time-invariance means that the HR to neuronal activity does not change with time.

Combining these two features, the system transferring neuronal activity into a HR is said to be a linear time-invariant (LTI) system. An LTI system transforms input to output via convolution with an impulse-response function, i.e. the response of the system to an input of *short* duration. In our context, the input being neuronal activity, this response is by definition the HR. Therefore, the BOLD response is mathematically described as the convolution of the stimulus pattern and the HR,

as indicated in Fig. 1.3b.

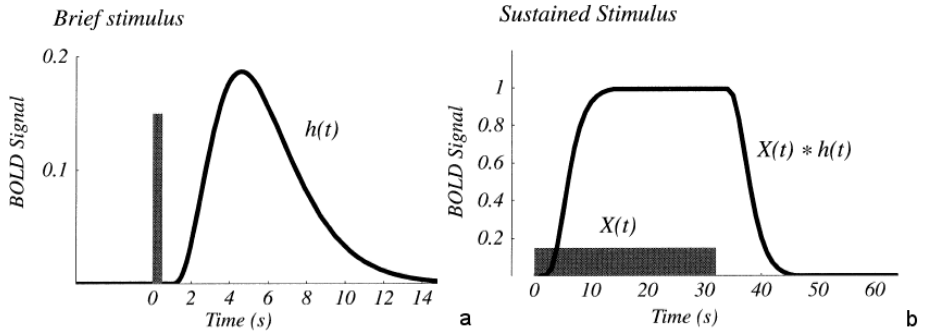

Fig. 1.3: *Model of the haemodynamic response to a brief stimulus (a) and the BOLD response to a sustained stimulus (b) [9].*

Figure 1.4 shows a realistic example of a BOLD signal response originating from a human visual stimulus experiment [52]. The response exists of different phases.

- After the initiation of the stimulus, there is an initial delay of 1-3 s [53], during which an initial dip may occur (as shown in Figure 4). This initial dip typically lasts 1-2 s [52, 54, 55] and reaches its maximum excursion of 1-2 % at 2-3 s after the stimulus onset for 4 T experiments [55], but is much weaker (compared to the late positive response) for 1.5 T experiments [56]. This fast response is potentially one of the most important aspects of the BOLD response (see section 1.7.2). However, the initial dip is not always seen, and this has led to controversy over its existence [57, 58].

- The initial delay (with or without initial dip) is followed by a ramp of about 5-8 s before a plateau signal change is reached.

- After the end of the stimulus the BOLD signal ramps down over several seconds and often undershoots the original baseline. Although the post-stimulus undershoot is not always evident, numerous examples can be found in literature [52, 55, 59]. Bandettini et al. reported undershoots with amplitude about half the plateau amplitude and that take 20 s to resolve [53]. Other authors reported more pronounced undershoots that take more than a minute to resolve [45, 58, 60].

1.4.3 Nonlinearity in BOLD response

The quantitative relationship between simple stimuli and the resulting BOLD response is examined by several investigators in experiments comparing the response

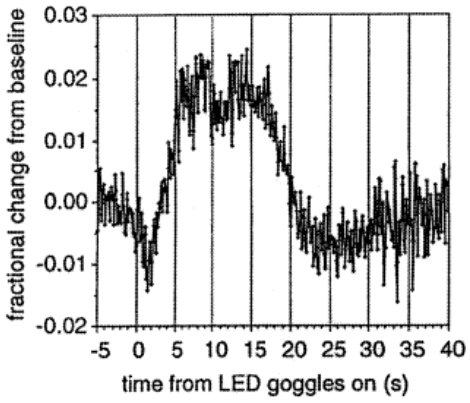

Fig. 1.4: *Example of a BOLD signal response originating from a human visual stimulus experiment [52].*

to brief stimuli to the response to longer stimuli. These studies used visual stimuli [14, 61, 62], auditory stimuli [63, 64], and motor tasks [63]. The consistent result of these studies is that, even though the response is roughly linear, there is a definite nonlinear component. The nature of this nonlinearity is that the response to a brief stimulus (e.g. $< 4\,\mathrm{s}$) appears stronger than would be expected given the response to a longer stimulus.

Possible explanations of this nonlinearity can be given by considering the process that leads from the stimulus to the BOLD response as consisting of three stages, as illustrated in Fig. 1.5 Nonlinearities may enter in each of these stages.

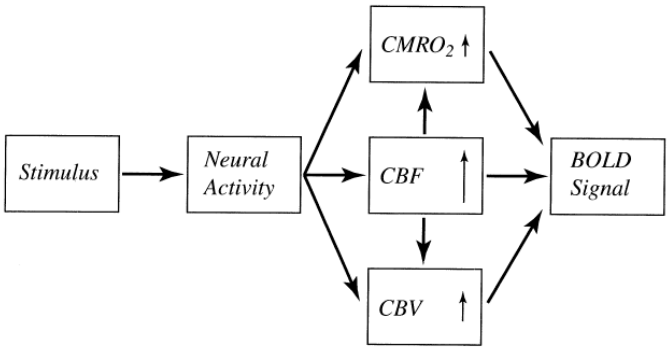

Fig. 1.5: *Different stages in the process from stimulus to BOLD response [9]. Nonlinearities may enter in each stage.*

(1) The first step is the translation of the stimulus pattern into a temporal sequence of neuronal activity. This neuronal response is likely to be nonlinear: it often begins with an initial peak of activity before settling down to a sustained plateau [65–67].

(2) The second step is the translation of the neuronal activity time course into haemodynamic changes (blood flow, blood volume, and oxygen metabolism) and is modelled as a simple linear convolution (see section 1.4.2).

(3) The third step is the translation of the haemodynamic changes into the BOLD response and is likely to involve two types of nonlinearity. Firstly, the metabolic changes follow different time courses, with the CBV lagging behind the CBF change (see section 1.3.4) [29, 47]. Secondly, the nature of the BOLD effect limits the signal to be measured in a BOLD experiment to a maximum, corresponding to full oxygenation. Thus any BOLD signal increase is an approach toward this ceiling. Any extrapolation of the BOLD change measured with a small flow change will overestimate the BOLD change for a large flow change.

Despite these nonlinearities, it is common in the analysis of BOLD data to assume linearity. This assumption undoubtedly introduces some error into the analysis [53], but many applications indicated that the error is likely to be small [14, 68].

1.5 Brain mapping by BOLD fMRI experiments

1.5.1 Visualising human brain activation

The influence of deoxyhemoglobin concentration in the blood on magnetic susceptibility was firstly confirmed by an MRI experiment performed at high magnetic fields (7 T and 8.4 T) on brains of live mice and rats [69]. The contrast in the acquired images depicted anatomical details of the brain by numerous dark lines, representing the blood vessels. A subsequent study consisting of a series of in vitro blood sample experiments and image simulations performed at high fields indicated that image contrast subsequent to local field variation around blood vessels was pronounced for gradient echo (GE) images and much weaker for spin echo (SE) images [70]. In an animal study, Ogawa et al. showed that blood oxygenation level dependent (BOLD) contrast follows blood oxygenation changes induced by anaesthetics and by inhaled gas mixtures that alter metabolic demand for blood [11].

These early animal studies in which blood oxygenation was manipulated by the experimenter suggested that BOLD contrast was suited to provide in vivo real-time maps of blood oxygenation in the brain under normal physiological conditions. The development of fast imaging techniques like Echo Planar Imaging (EPI), originally proposed by Peter Mansfield [71], offered the necessary improvement in time resolution as to measure these BOLD signal changes (i.e. time courses). An image consisting of a 128×128 matrix for example, can be measured using a GE pulse

sequence in less than 1 s. With EPI, this image can be obtained within a few tens of milliseconds.

Changes in image contrast were firstly visualized using EPI in a study on a cat model imposing respiratory challenges [36]. Kwong et al. were the first to map human brain activation using BOLD contrast, using visual and motor stimulation paradigms [12]. This report marked the beginning of functional human brain mapping experiments based on the BOLD effect.

1.5.2 Mapping brain functionality

Numerous studies have demonstrated that the human brain is segmented into many distinct areas that are functionally specialized. However, the spatial scale of this functional specialization varies. Although very interesting, the detailed description of the latter specialization is considered beyond the scope of this dissertation and is therefore omitted. An excellent description can be found in [72]. For the determination of the location of the various anatomical structures, we refer to the human brain atlas of Talairach and Tournoux [73]. In the following paragraph, a general review of literature concerning functional specialization is presented.

Basic functions Some brain functions, like speech and language [74–76] are part of a distributed network located primarily within the left hemisphere. Other functions, such as vision [77–85], somatosensory [86], and motor control [87–90] have their cortical representations located within both hemispheres, and are localised within gyri and sulci tens of millimetres in size.

Auditory system The study of the auditory system [91, 92] proceeded considerable slower compared to other functional systems, due to several factors. The first concerns the intrinsic anatomy of the auditory system, i.e. the reduced extension in space and the inter-individual variability of regions involved in auditory perception. The second concerns specific difficulties arising mainly from the loud acoustic noise, produced by the gradient system during image acquisition. This considerable background noise can interfere with the experimental stimuli in an unpredictable way. Among the many effects derived from this interference, the most expected one occurs in the partial saturation of neuronal-evoked regional haemodynamics that reduces, probably in a non-linear fashion, the BOLD signal amplitude in response to the auditory stimuli. To overcome these problems, different approaches have been proposed, which generally require a careful tailoring of the experimental designs, the fMRI methodology, and the strategies of data processing to the specific problems of audition and the particular research goal. An excellent review of solutions already existing or under development to the specific problems of auditory system, like 'silent fMRI', can be found in [93].

Complex tasks In the subsequent years, brain areas related to more complex motor and sensory tasks or cognitive tasks were investigated. Studies were for

instance performed in the field of word generation [94], puzzle solving [95], motor skill-learning [96], language-related brain organization [97], or border determination of multiple visual areas [98–101]. In the latter studies, the visual cortex was found to be organised into multiple specialised areas, namely areas V1 to V8.

Mental imagery Another vast area of research is mental imagery, which occurs when perceptual information is assessed from memory. By contrast, perception occurs when information is registered directly form the senses. Mental images do not only result from the recall of previously perceived objects or events, they can also be created by combining and modifying stored perceptual information in novel ways.

Among the best studied mental imagery tasks are Paivio's 'mental clock' task [102] and Shepard and Metzler's 'mental rotation' task [103]. Both tasks require several seconds of processing time and thus are particularly well suited for an fMRI-based mental chronometry investigation. In the mental rotation task photographs of two rotated objects are compared and subjects have to report whether the two objects are identical or mirror-inversed. In the mental clock task, pairs of clock faces are compared on the basis of acoustical presented times. Subjects have to compare the mental images and report in which of the two faces the clock hands form the greater angle.

Research concerning mental imagery can be generally divided in three classes. A first class of studies performed visual [104–108], auditory [109, 110], and motor [111–113] mental imagery experiments to collect evidence that imagery engages brain mechanisms that are used in perception and action. A second class of studies illustrated that visual mental imagery engages even the earliest visual cortex. Some of the original areas have been further subdivided. [114–116]. Finally, a last class showed that imagery engages mechanisms that control physiological processes such as hart rate and breathing, having effects much like those that occur with the corresponding perceptual stimuli [117]. An excellent review of neural foundations of imagery and corresponding neuroimaging studies is found in [118].

1.5.3 Block versus event-related paradigms

Basically two types of experimental designs for the localization of human brain functionality exist: block and event-related. Due to specific shortcomings related to block paradigms, a shift to event-related paradigms occurred.

Block paradigms

The first fMRI studies used a 'block' design. Early block design experiments alternated 'ON' periods of activation, corresponding to the execution of a task, with

'OFF' control periods of rest. Rest, however, is defined only with respect to the specific activity being considered and contains ongoing baseline activity. Generally however, block designs consists of different tasks (instead of task-rest) and brain responses to these tasks during the experiment are acquired and compared. Each task is of roughly the same duration, typically in the range 20-30 s. The alternation ensures that signal variations from small changes in scanner sensitivity, patient movement, or changes in attention have a similar impact on the signal responses associated with each of the different states. As the shape of the BOLD response function can assumed to be simple, regions of activity change between one condition and another can be identified with considerable statistical power (see section 2.2.1). The MR acquisition methods used in conjunction with block paradigms are indeed designed to give both maximal SNR and maximum BOLD contrast for whole brain coverage. An extensive overview of possible designs with blocked paradigms can be found in [119, 120].

However, block design creates a highly artificial psychological constraint: it may be difficult to control a cognitive state precisely for the relatively long periods of each block. A 'rest' state is rarely a true rest, as the mind 'wanders' in a subject who is not engaged in a specific task. Many types of stimuli (particularly sensory stimuli) may show rapid habituation. The latter phenomenon is defined as the reduction of responsiveness after prolonged or repeated exposure to a stimulus. It is an adaptive ability to cease responding to irrelevant events in an environment with multiple sensory stimuli [121]. More complex cognitive tasks simply may not be amenable to a block design. Finally, information regarding the time course of an individual response is lost within a block.

Event-related paradigms

To overcome these shortcomings, the so-called 'event-related' (ER) design was presented [122, 123]. Event-related paradigms differ from blocked paradigms in that individual trial events are measured, rather than a temporally integrated signal. Such a design may seem counterintuitive given the lag and the temporally blurring inherent in the HR. However, the response within a given subject and within a given region of cortex was found to be very consistent from one set of measurements to the next [61, 124, 125]. The lag of onset and the time course of signal evolution are highly reproducible. Activity can thus be revealed by averaging data acquired after many such discrete events. Averaging increases the SNR for signal changes when activity is time-locked to the experimental event, i.e. when the activity is aligned with the event of interest [122, 126].

Several features of fMRI data proved to be critical in allowing event-related procedures to be developed.

- Firstly, technological advances in fMRI data acquisition speed allowed acquiring the data over the time course of an individual event. Techniques like EPI acquire data extremely rapidly: if only a few slices through the brain

are acquired, measurements can be repeated in less than 1s, whereas whole brain coverage can be acquired in as little as 2 s. The effective sampling rate can even further be reduced by appropriate paradigm design. Miezin et al. for example doubled the effective sampling rate using an interleaved procedure [125]. For a fixed TR of 2 s, two data sets were acquired providing odd and even sample points that contribute to a composite waveform with 1 s sampling resolution.

- Secondly, even for very brief periods of neuronal activity, resulting signal changes were measurable despite the delayed and prolonged nature of the time course of the HR [53, 80, 127], indicating high fMRI signal sensitivity. Savoy et al. demonstrated that a visual stimulation as brief as 34 ms in duration will elicit small (i.e. signal changes less than 1%) but clearly detectable, signal changes, as shown in Fig. 1.6 [128].

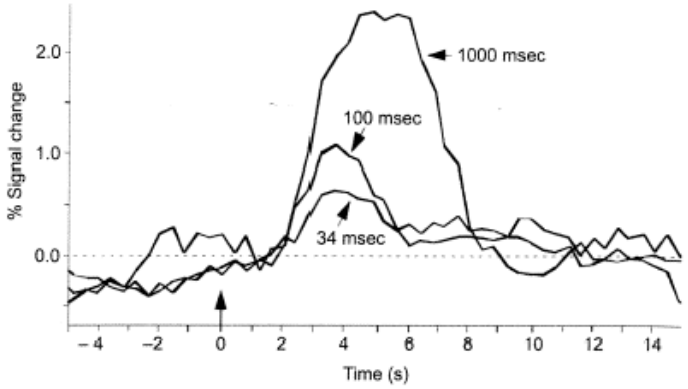

Fig. 1.6: *Sensitivity of the BOLD signal. Even very brief periods of neuronal activity (as brief as 34 ms), result in measurable signal changes, despite the delayed and prolonged nature of the time course of the HR [128].*

- Thirdly, the HR to neuronal activity has been shown to provide a highly consistent response that summates over sequential events in a roughly linear fashion (see above). In the event-related context, linearity means that the shape of the BOLD HR to a given period of stimulation is predictable and relatively stable across events, even when there is an overlap in the response to successive events, as illustrated in Fig. 1.7 [14, 61]. However, the full impact of these nonlinearities, particularly in event-related experimental paradigms that involve the separation of overlapping responses, has not been explored until now.

Finally, we mention that, more recently, mixed designs were introduced. They provide the power and flexibility to address more complex cognitive hypotheses that are otherwise difficult to study with block design or event-related design alone.

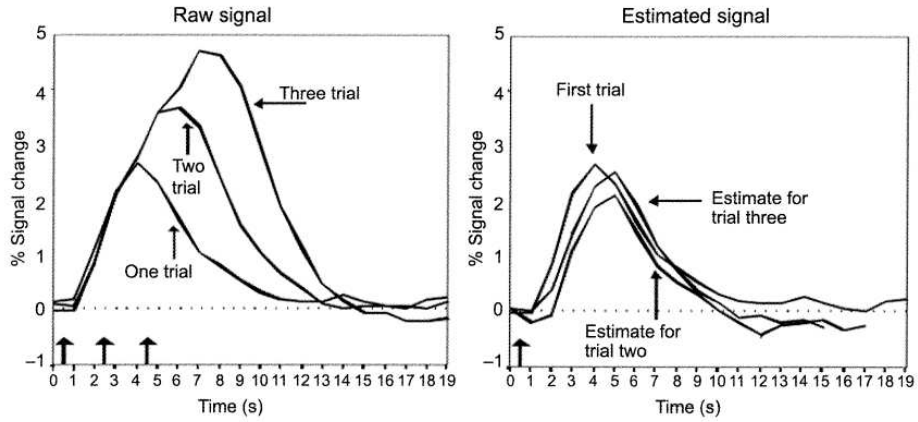

Fig. 1.7: *Event-related fMRI data show approximately linear summation of the HR for closely spaced trials. The left figure shows the fMRI signal intensity elicited by the presentation of one, two, or three trials. The response is increased and prolonged by the addition of multiple trials, but does not saturate as successive trials overlap. On the right, the estimates of the individual responses to each of the three trials are shown [61].*

Laurenti et al, for example, superimposed in time visual blocked stimuli and auditory events [129]. They demonstrated, using a GLM procedure, that regional activity corresponding to both stimuli could be separated.

1.6 Non consistently-task-related BOLD effects

Neuroimaging literature is predominantly concerned with the determination of the precise location and time course of consistently-task-related (CTR) responses corresponding to true neuronal activity. BOLD fMRI tries to measure the neuronal activity indirectly by means of the haemodynamic effect. The latter however is associated with activation through a neurovascular coupling composed of a cascade of complex and poorly understood mechanisms. Several non task-related effects could occur during the linear transform pathway, leading to signal increases that masquerade as BOLD activations. Various sources of noise as well as sources of non-neuronal origin, such as the inflow effect, contribute to the fMRI response. Other confounding effects are neuronal in origin, but do not reflect consistently-task-related activity, like transiently-task-related (TTR) activity, default activity, or negative BOLD response.

Several approaches exist to deal with these effects. An appropriate experimental design is used to prevent their occurrence as much as possible. Additionally,

after data collection, remaining effects can be somewhat corrected with preprocessing techniques. Finally, a careful interpretation of the activation patterns found by the applied analysis method is performed post-hoc. Remaining artefactual components can thus as yet be identified and subsequently excluded from further analysis. In this section, we will elaborate on the origin of the confounding effects as well as on specific techniques to prevent subsequent corruption of the analysis results.

1.6.1 Noise

In the fMRI context, noise is (broadly) defined as the nuisance signal that disturbs the fMRI response. Confounding effects are introduced at each stage of the linear transform pathway for the transformation of stimulus related neuronal activity into a BOLD response. However, according to the linear transform model of Boynton et al. (elaborated for a visual stimulus), the effect of these individual noise sources can be summarized by a single noise source that adds to the desired fMRI signal, provided that the sources are independent of stimulus and temporal period [14] (see Fig. 1.8).

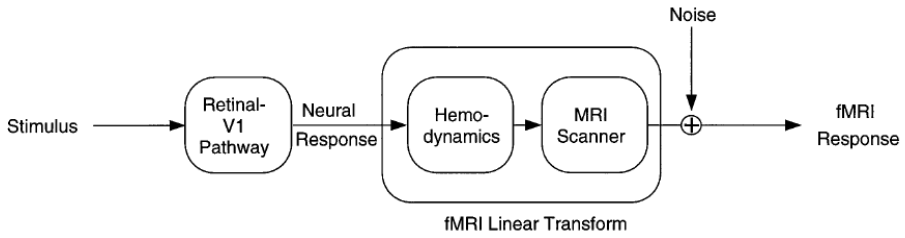

Fig. 1.8: *According to the linear transform model of Boynton et al, the effect of individual noise sources can be summarized by a single noise source. [14]*

In order to determine techniques to reduce the noise-component in the measurement data, noise sources are classified based on their characteristics. The latter approach yields two categories: thermal noise and physiological noise.

1.6.1.1 Thermal noise

Characteristics

Thermal noise refers to thermal fluctuations in the electronics of the scanner hardware and the detector coil on the one hand and fluctuating magnetic fields originating from random (stray) currents in the body on the other hand. Thermal noise is independent and identically distributed (i.i.d.) and can accurately be described by a Gaussian distribution. However, in practice the images acquired by

the scanner are submitted to several transformations, influencing the aforementioned characteristics of the noise component. Originally, the scanner acquires a complex (i.e. consisting of a real and imaginary component) image in the so-called *k-space*. For a description of image acquisition and mapping in k-space, we refer to [9], chapter 10. An inverse Fourier transform results in a *spatial* image. Due to the orthogonality of the Fourier transform, Gaussian distributed k-space noise will still be Gaussian distributed in the spatial domain. Spatial image data remain complex valued. This complex image is generally transformed in a *magnitude* image and a corresponding phase image. The phase image is in fact seldom used by the current analysis methods (and often even not available). The thermal noise component which is characterised by a Gaussian distribution in the complex image, now becomes Ricean distributed in the magnitude image [130]. However, a Ricean distribution may be approximated by a Gaussian distribution for an SNR (see below) larger than 10. Since most fMRI images exhibit an intrinsic SNR on the order of 100, the latter condition is met. Therefore, thermal noise in magnitude data may be described by a Gaussian distribution. The noise parameters can be estimated from MR data as to optimally describe the underlying Gaussian distribution. [131–133].

Signal-to-noise ratio

Several studies pointed out that the ratio of the signal amplitude to the noise standard deviation is a critical factor to determine whether an acquired MRI signal can be used to extract relevant information [134–136]. As already mentioned, this quantity is called the signal-to-noise ratio (SNR) of an image. This *image* SNR depends both on the *system* SNR and the *intrinsic* SNR.

- The system SNR reflects the influence of the hardware, such as the coils. In practice, it is determined by a trade-off between small coils for better SNR and large coils for better coverage. Indeed, a small coil picks up less stray noise from the rest of the body, but provides only limited coverage of the brain. This trade-off can be overcome however by using multiple small coils in a phased array system [137]. The latter solution is at the expense of the uniformity of the sensitivity pattern and requires a scanner hardware configuration with multiple receiver channels.

- The intrinsic SNR is determined by the intrinsic properties of the measured sample. It is described as the signal from the sample volume competing with thermal noise in the sample (i.e. body noise). Physiological noise (see below) is not an intrinsic property of the measured volume and is therefore not reflected in the SNR definition. The intrinsic SNR can be quite large. In a single-shot EPI-image it is on the order of 100. It is affected by three factors:

$$\mathrm{SNR} = M \Delta V \sqrt{T_{\mathrm{read}}} \quad . \tag{1.2}$$

The first factor (M) is the intrinsic magnitude of the generated transverse magnetization $\boldsymbol{M}$ at the time of the data acquisition. It is expressed in terms of the magnetic field strength $\boldsymbol{B}_0$ and parameters depending both on tissue and pulse sequence. The second factor is the voxel volume (ΔV) while the third factor (T_{read}) is the total duration of the data acquisition, the so-called 'readout'. A detailed description of all influencing parameters can be found in [138], pages 112-113.

Contrast-to-noise ratio

At first glance, it might appear that optimizing the SNR is preferred. However, maximizing the SNR is not optimal for anatomical nor for functional imaging. Indeed, instead of the signal amplitude, the contrast is what determines whether one tissue can be distinguished from another in the image (MRI) or whether an activated region can be distinguished from a non-activated one (fMRI). The latter concepts are called anatomical or within image contrast (MRI) and functional or between image contrast (fMRI), respectively. A simple difference map of the contrasts however, would be dominated by vessel and CSF artefacts. Indeed, due to the non-uniform physiological noise component, voxels with the largest fluctuations will show the largest difference signal. Therefore, in most fMRI-applications, the ability to distinguish between activated and non-activated regions depends not just on the contrast, but rather on the 'contrast-to-noise ratio' (CNR) [139, 140]. The latter ratio 'normalizes' the contrast by dividing it by an estimate of the intrinsic variability of the signal in the voxel.

Removing thermal noise

Consistently task-related BOLD signal change during activation is quite small, on the order of 1 % for a 50 % cerebral blood flow change at 1.5 T. The intrinsic SNR in a single-shot EPI-image on the other hand is on the order of 100. Although this is quite large, it results in a poor CNR, not much larger than 1. This is not large enough to reliably detect this BOLD signal change in a voxel from a single image in the stimulus state and a single image in the control state. For this reason several approaches are applied to deal with thermal noise.

Averaging A first approach is to acquire a large number of images to allow sufficient averaging over trials. For random signal variations with a normal distribution and independence for each time point, averaging will improve the SNR proportional to the square root of the number of averages. The CNR is approximately linearly related with the intrinsic SNR (see 1.7.2).
Although averaging the data over several trials improves the detection sensitivity of CTR signal changes, it must be applied with care. Indeed, averaging reduces the detection sensitivity for TTR signal changes in the fMRI signal. Moreover, averaging reduces sensitivity to CTR changes not consistently time-locked to the experiment design. These CTR changes typically include

changes in strategy by the subject during the test period, changes associated with learning or habituation of task performance, with fatigue, or with other processes whose time course cannot be predicted in advance by the experimenter.

Smoothing Thermal noise can also be reduced by temporal or spatial smoothing. Both filtering approaches are subject to constraints however. The width and type of the spatial filter are chosen on the basis of a trade-off between the desired spatial resolution and the expected enhancement of the functional CNR. In case of temporal smoothing, the choice of the bandwidth of the filter is driven by a trade-off between the expected enhancement of functional CNR and the loss of temporal resolution [141, 142].

Statistical approach As an alternative, several statistical approaches are applied to deal with the Gaussian noise component in the signal. Descombes et al. used 'Markov Random Field' theory to account for the non-independence of neighbouring voxels [143]. Task-induced neuronal activations and BOLD responses are indeed expected to produce similar signal changes in spatially contiguous regions, extending over several millimetres. Therefore, Descombes et al. proposed to use this additional information in the detection process and defined some interaction between neighbouring voxels which allowed reducing the noise while preserving the signal characteristics. Ardekani et al. modelled the nuisance component representing effects of no interest and Gaussian white noise using a maximum likelihood procedure [144].

Acquisition-based approach Techniques mentioned so far to reduce thermal noise in a voxel were all based on *preprocessing* approaches. Alternatively, *acquisition-based* approaches exist. They focus on decreasing the bandwidth, increasing the field of view (while maintaining the resolution), and increasing the voxel size (to increase the signal amplitude rather than to decrease the noise). As these parameters are fixed in the experiment providing the measurement data used in this dissertation, we will not further elaborate on their use.

1.6.1.2 Physiological noise

Characteristics

Additionally to thermal noise, the BOLD signal is corrupted by physiological fluctuations, originating from different sources:

- Cardiac pulsation (on the order of 1 Hz) creates a pressure wave that strongly affects the signal of flowing blood, but it also creates pulsations in CSF and in the brain parenchyma itself.

- Respiratory motion has a period of several seconds and changes the shape of the body. It exhibits long temporal correlations and broad spatial patterns.

- Vasomotion is a regular oscillation of blood oxygenation and flow due to constriction or dilation of blood vessels [145]. It contributes to low-frequency oscillations of the BOLD signal, particularly at high magnetic fields.

- In its broadest sense, physiological noise also includes subject movement. If there are edges in the intensity pattern of the image, such as near the edge of the brain, movements much smaller than a voxel dimension may produce a signal change larger than the expected signal change due to the BOLD effect. This effect is particularly troublesome if the motion is correlated with the stimulus [146], e.g. if the subject tips his head slightly every time a visual stimulus is presented or a button is pressed. The effect can create large apparent activations that are purely artefactual.

Although most physiological motions are small, they nevertheless all can produce signal fluctuations of the same order as the BOLD effect. The variance of the signal over time measured in a human brain is several times larger than would be suspected from thermal noise alone. Furthermore, physiological noise is likely to violate assumptions of normal distribution and independence, as it exhibits both temporal and spatial structure [147, 148].

Removing physiological noise

There are several approaches dealing with physiological artefacts.

- Motion is best prevented as much as possible in advance. This is accomplished by carefully coaching the subject about the importance of remaining still. Stimulus-correlated motion can be considerably prevented by using head restraints. Low-frequency oscillations [149] in the acquired data, including heartbeat- and breathing-related motion, can be eliminated by temporal filtering techniques. Motion-correcting algorithms can be applied to produce the best mutual alignment of the images. In a recent study, Oakes et al. performed a comparison between several motion correction software tools [150].

- Confounding effects related to subcortical structures, such as the ventricles and the white matter are eliminated by restricting the analysis to effects originating in the cortex (i.e. the grey matter). The latter is accomplished by a preliminary segmentation of the cortex [151].

- In case of thermal noise, averaging improves the SNR proportional to the square root of the number of averages and the CNR approximately proportional to the latter (see 1.7.2). For physiological noise, which is neither normally distributed nor uncorrelated, this factor will be reduced.

- Several authors investigated alternative techniques to deal with physiological noise. Biswal et al. used digital filtering techniques [152]. Frank et al. proposed a Bayesian model which proved to be particularly suited to handle

the spatial variations in the noise present in fMRI, allowing the comparison of activated voxels that have different and unknown noise [153]. Restom et al. proposed techniques within the GLM framework to reduce physiological noise for ASL fMRI [154]. Triantafyllou et al. investigated the effect of spatial smoothing on physiological noise [155]. They found that at 7 T, $5 \times 5 \times 3$ mm^3 resolution images derived from smoothing $1.5 \times 1.5 \times 3$ mm^3 data improved time course SNR by a factor of 1.89 compared to a time series acquired at $5 \times 5 \times 3$ mm^3. They presumed that this effect was derived from the reduced physiological-to-thermal noise ratio in the high spatial resolution data followed by a smoothing operation that improves SNR without adding physiological noise.

1.6.2 Neuronal and haemodynamic effects

Transient effects

In block design fMRI experiments, a number of transient patterns were reported to occur at the transitions between rest and active state. These dynamic aspects include signal overshoot and undershoot at both beginning and end of the stimuli. Transients in the BOLD response could be an accurate reflection of transients in the neuronal activity itself. However, because the BOLD signal depends on the combined changes in CBF, CBV, and $CMRO_2$, such transients also can arise due to mismatches in the timing of the metabolic changes (see section 1.3.4).

In order to test the origin of the transients of the BOLD response, several authors measured CBF directly with ASL techniques. The question was not unambiguously resolved however, obliging cautiousness when interpreting transient features of the BOLD signal, especially without measuring the flow response. Hoge et al. found that initial overshoots and post-stimulus undershoots represented the temporal pattern of neuronal activity rather than time lags of physiological changes [35]. However, in a study combining motor and visual activation, Buxton et al. found a pronounced post-stimulus undershoot in the BOLD signal without corresponding post-stimulus undershoot in the flow signal [156]. Other authors had similar findings [47, 157].

Default activity

Functional brain image studies in normal human subjects with PET and fMRI have frequently revealed task-induced decreases in regional brain activity that appear to be largely task-independent, varying little in their location across a wide range of tasks [158]. The consistency with which certain areas of the brain participate in these decreases led Raichle et al. to posit the existence of an organised mode of brain function that is present as a baseline or default state and that is at-

tenuated or suspended during specific goal-directed behaviours [159]. The default activity might reflect ongoing processes, such as unconstrained verbally mediated thoughts and monitoring of the external environment, body, and emotional state.

Negative BOLD signal changes

Negative BOLD responses refer to *sustained* negative responses to a stimulus that causes positive activation elsewhere. We emphasize that the undershoot of the baseline on cessation of stimulation and the small initial dip in the BOLD response are no sustained responses and therefore do not refer to a negative BOLD response.

Several examples of negative BOLD were reported. In a visual stimulation study, Smith et al. noticed that the resulting region of activation (positive BOLD) in the primary visual cortex was surrounded by an extensive area of deactivation (negative BOLD). The latter area affected much of the visual cortex, apart from the activated area [160]. The phenomenon of negative BOLD in the occipital cortex has been noted by various other authors [161–163] and is also seen with PET [158]. There are both PET and fMRI studies of deactivation of a sensory or cognitive system in association with activation of another system, although it is not clear whether this reflects similar processes to negative BOLD in unstimulated portions of the visual cortex [158, 164, 165].

Although the occurrence of negative BOLD in occipital cortex is not in doubt, its interpretation is controversial. Several hypotheses exist.

- Raichle et al suggested that it reflects 'inhibition' or 'suppression' of neuronal activity, which causes an actively controlled reduction in CBF [166]. This was further supported by Shmuel et al [167].

- However in the latter study, Shmuel et al. raised the possibility of a contribution of haemodynamic origin. They proposed that the negative effect might also reflect 'blood stealing'. The elevation of blood flow at the activated location could cause pressure changes which reduce blood supply in nearby areas sharing the same blood vasculature (without compensatory decrease in $CMRO_2$). The suggestion of reallocation of cortical blood resources was supported by Harel et al. [168].

- Smith et al. provided evidence against 'blood stealing' as the mechanism for this reallocation and proposed the occurrence of another driving mechanism, called 'blood redistribution' '[169]. Instead of passively stealing blood by local pressure changes, the flow could be actively restricted elsewhere, by neurally controlled constriction of vessels.

- Finally, the negative BOLD could also be the result of an increase of $CMRO_2$

without a compensatory increase of CBF, much as in the initial dip phenomenon.

Inflow effect

Inflow effects refer to the non-saturated spins of blood flowing into the excited slice. The inflowing spins are not saturated because they have not experienced any RF-pulses yet. However, when the spins move into the slice, they are excited by RF-pulses and become partially saturated. During activation of a brain region blood flow increases, resulting in an increase of the blood velocity.
Gomiscek et al. showed that if the mean flow velocity in the vessels is lower than a critical value (depending on the sequence parameters and tissue), the partial saturation and, consequently, the signal intensity in the vessel depends on blood velocity [170]. They illustrated that an increase in the blood velocity results in a decreased partial saturation of spins and an increase in signal intensity in the voxel containing the vessel (signal intensity increases of 10-30 % were measured).

The inflow effect yields a localisation problem for BOLD sequences. Blood flow increases to the cortical region of increased brain activity are regulated by arterial vessels, that can be a centimetre or more away from the active site. Arterial inflow effects have been demonstrated to contaminate BOLD fMRI images under rapid RF pulsing conditions [95, 171, 172]. Due to rapid RF pulsing, most fMRI is not done in a fully T_1 relaxed MR state. Therefore, the potential exists that the inflow effects in these arterioles may result in signal increases that masquerade as activation changes but which are not co-localised to the neuronal activity.

Several techniques exist to remove the effect, like outer-volume RF saturation [171] or velocity nulling (also called diffusion weighting) [173, 174]. The latter technique applies a bipolar gradient pulse to selectively suppress the intravascular component with only a small effect on the extravascular signal [175]. As a result large vessel effects in GE and SE fMRI are suppressed as illustrated in Fig. 1.9 and Fig. 1.10.

1.7 Specificity of the BOLD effect

1.7.1 Spatial relationship between BOLD response and neuronal activity

Reliability of BOLD response

The reliability of the BOLD signal as an indirect measure of neuronal activity has received support from the comparison of the results of numerous fMRI experiments with other techniques (like a direct electrode recording [176–178]) as well as from the well-established body of literature on the functional organisation of the brain. They all clearly suggest that BOLD signal changes reflect some aspect of

neuronal activity. Several authors also demonstrated reproducible spatial patterns of activity associated with a task in a comparison across studies [179–181].
A recent review describing the current understanding of the causal relationship between neural activity and the BOLD signal and how this challenges some basic assumptions that have guided neuroscience can be found in [182].

Consistency of BOLD response

Many authors found that the BOLD response within a given subject and within a given region of cortex was extremely consistent from one set of measurements to the next [61, 125, 126, 183]. Therefore, although the signal is temporally blurred, the lag of onset and time course of signal evolution are highly reproducible.

The response of a given cortex region between subject exhibits a larger variability. Moreover, the HR varies across brain regions within a given subject. This variability is exhibited in terms of the onset and shape of the response, as indicated by several studies.
Data from Buckner et al., for example, revealed the onset of the response in secondary visual cortex to be 1-2 s earlier than the response in prefrontal cortex during a word generation task [184]. A related finding was presented by Schacter et al. in the context of a memory study, revealing a difference of several seconds between activity in anterior and dorsal prefrontal cortex [185]. Even within visual cortex itself, Bandettini showed variance in the timing of responses of 1-2 s across voxels of different subregions [127]. Patterns of activity within primary and secondary visual cortex could be predicted on the basis of the well-understood retinotopic organisation of the primary visual cortex [68, 98, 99]. Retinotopy reflects the principle that greater precision regarding execution of a task requires the involvement of a greater cortical surface.

Intra- versus extravascular signal changes

The BOLD signal consists in an intravascular component, referring to the blood vessels, and an extravascular component, referring to the brain tissue. Brain activation is located in the tissue. Since these components cannot be separated out of the measured BOLD signal, the accuracy in mapping an activation volume in brain tissue depends, amongst others, on how large the extravascular component is compared to the intravascular one.

For GE BOLD experiments however, for a field strength of 1.5 T and for typical voxel volumes of 30 mm^3, the contributions of the intravascular and the extravascular component are comparable [175]. The latter results from two opposite effects. The dynamic range for the intravascular signal change is more than an order of magnitude larger compared to the extravascular one, on the one hand. Within the blood, large field gradients are produced around the red blood cells carrying the deoxyhaemoglobin [186] so that at rest the venous blood signal may be reduced by

as much as 50 % compared to what it would be if the blood were fully oxygenated [175]. On the other hand, however, the intravascular compartment, is only a small fraction (about 4 %) of the total tissue volume (i.e. cortical grey matter). The latter reduces the influence of the first effect.

Micro- versus macrovascular signal changes

The BOLD signal consists in a microvascular component, referring to the capillaries, and an macrovascular component, referring to the veins. Brain activation is located near the capillaries. Since these components cannot be separated out of the measured BOLD signal, the accuracy in mapping an activation volume in brain tissue depends, amongst others, on how large the microvascular component is compared to the macrovascular one.

Gradient echo BOLD experiments are more sensitive to the veins than to the capillaries, a phenomenon resulting from two competing effects [81]. Due to the nature of the haemodynamic effects, the range of variation of the haemoglobin saturation is reduced in capillaries as opposed to the veins, on the one hand. On the other hand, signal reduction due to diffusion is different between micro- and macrovascular components. Diffusion effects are indeed small in veins because their diameter is large compared to a typical diffusion distance (as opposed to capillaries). As a result of these two effects, the largest BOLD signal changes are likely to occur around draining veins [187]. The latter may be removed from the area of neuronal activation, so the location of the BOLD change could differ by as much of a centimetre or more from the area of increased neuronal activity. Arterial spin labelling (ASL) experiments indeed show that the locations of the largest CBF change and the locations of the largest BOLD signal do not always coincide [188, 189]. Fig. 1.9 illustrates the signal seen from the vessels and the surrounding tissue for the GE pulse sequence (velocity nulling: see section 1.6.2).

1.7.2 Improving the specificity

Vast research is performed concerning the improvement of the specificity of the BOLD signal. We expound several factors of influence.

Short duration stimuli

The use of short duration stimuli prevents the BOLD response from saturating and permits the use of peak fMRI signal magnitude as a measure of the amount of neuronal activity within a voxel [14].

Several authors reported poor functional resolution capabilities of fMRI when trying to increase SNR using a block design experiment [68, 190]. In visual task experiments using stimuli of at least 10 s in duration, they found that, when the

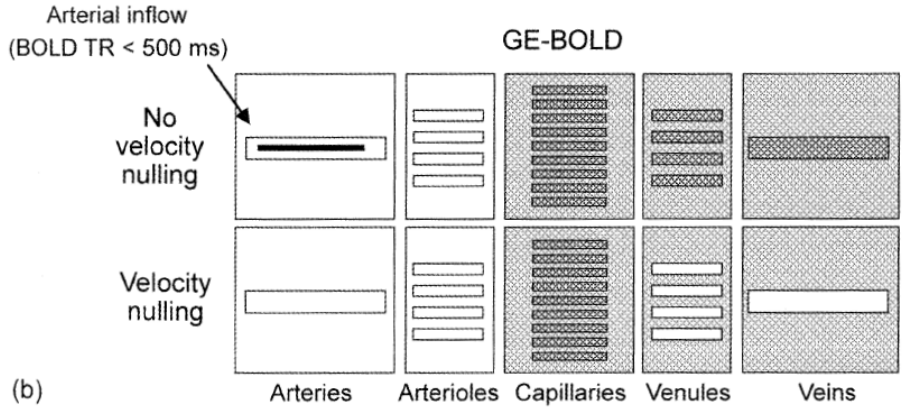

Fig. 1.9: *Localisation of BOLD-signal (grey colour) for a GE pulse sequence. The small rectangles represent the vessels, whereas the larger rectangles reflect the surrounding tissue. 'Velocity nulling' suppresses the section 1.6.2 [138].*

BOLD response reaches a plateau, the measured fMRI signal extends to areas well beyond the site of the excited neuronal tissue. This makes functional resolution at a submillimetre scale unlikely for block design experiments. In designs using brief stimuli (less than 4 s), the locus of neuronal activity may still be resolvable, at least from adjacent areas that are activated under an orthogonal stimulus condition.

High field strength

Gati et al. investigated the influence of higher field strengths, i.e. 3 T or 4 T, in a visual experiment study [191]. They presented several findings.

- Microvascular contributions increased relatively compared to macrovascular contributions for higher field strengths.
- The average percentage signal change accompanying the increasing field strength was larger in vessels compared to tissue, however.
- The SNR in fully relaxed proton density weighted GE images was found to increase linearly with respect to the magnetic field strength.
- The CNR was found to behave less than linearly with field strength in voxels containing vessels larger than the voxel itself and greater than linearly with field strength in voxels containing a mixture of capillaries and veins with a diameter less than the voxel.

The optimal field strength for performing BOLD-based fMRI experiments has been matter of debate however [192, 193]. Indeed, the actual choice of the scanner's magnetic field strength depends on several trade-offs, which include not only the aforementioned aspects, but also bandwidth considerations [194], RF power deposition [195], acoustic noise of the gradients [196], and physiological noise constraints [197, 198]. Especially the latter seems to impose a limit on further gains in SNR, because physiological noise becomes a larger fraction of the total noise when the signal increases due to higher field strength.

SE pulse sequence

Higher field strengths benefit the SE pulse sequence compared to the GE pulse sequence in terms of specificity, without constraints imposed by the sensitivity. In order to illustrate the latter, we elaborate on the sensitivity and specificity of SE and GE and the influence of the field strength on these characteristics.

sensitivity Typically, in the first fMRI studies with field strengths of 1.5 T, GE pulse sequences were applied. The sensitivity and the corresponding image contrast subsequent to local field variation around blood vessels were pronounced for GE images. For field strengths of 1.5 T, SE sensitivity was too small for accurate detection of brain activity [70]. At higher field strengths (4.7 T to 9.4 T) however, the sensitivity of a SE sequence no longer hampers an accurate detection of brain activation.

specificity For the comparison of SE and GE in terms of specificity, both the contribution of the extravascular versus intravasculare component and the contribution of the microvascular versus the macrovascular component is considered. These contributions are illustrated in Fig. 1.9 and Fig. 1.10 for GE and SE, respectively (velocity nulling: see section 1.6.2).

- The extravascular SE signal change at 1.5 T is more sensitive to the smallest vessels, like the capillaries and small venules, as opposed to GE.
- The intravascular SE signal change at 1.5 T is dominated by the veins [199, 200]. The macrovascular component is thus larger than the microvascular one for SE. However, at higher field strengths, Lee et al. estimated that the fractional contribution of the venous blood signal change to the net SE BOLD signal change is 60 % at 1.5 T, 8 % at 4.7 T, and 1 % at 9.4 T [201].

Optimal voxel size

The determination of the optimal voxel size, i.e. spatial resolution is submitted to particular constraints. When high spatial resolution is employed, the resulting

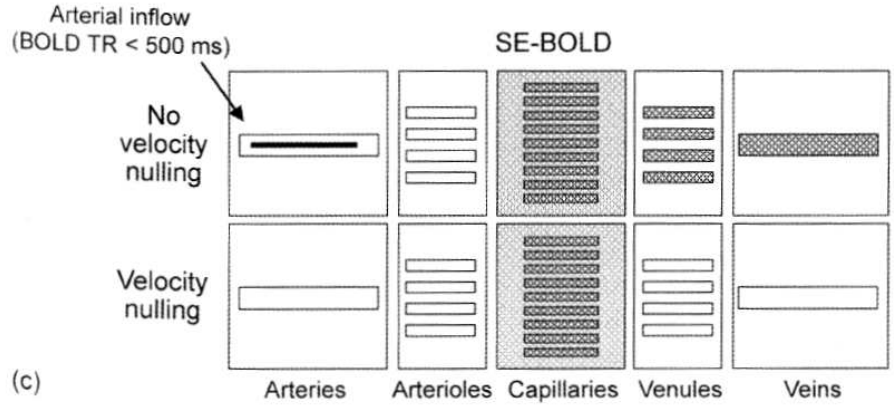

Fig. 1.10: *Localisation of BOLD-signal (grey colour) for a SE pulse sequence. The small rectangles represent the vessels, whereas the larger rectangles reflect the surrounding tissue. 'Velocity nulling' suppresses the section 1.6.2 [138].*

drop in SNR and CNR implies that only voxels exhibiting a large change in signal will be detectable. Large vessels will thus dominate high resolution BOLD maps of brain functioning. Lower spatial resolution on the other hand deteriorates the functional localisation due to the partial volume effect. Since functional region sizes are not well defined and are likely to vary over space and experimental question, the optimal voxel size, corresponding to a trade-off between both influences, is difficult to predict. Several approaches exist.

- Matching the voxel volume to the cortical thickness, which is about 3 mm, is a common practice. Therefore, in many BOLD experiments the voxel size is typically about 30 mm^3, leading to BOLD activations of a few percent at 1.5 T.

- Frahm et al. noticed that, when the voxel size of the images is reduced, the amplitude of the largest BOLD signal changes increases dramatically (to 20 % and larger), suggesting that the changes are localised to a region smaller than 1-2 mm [202]. Recent work suggests that the optimal volume dimensions are $1.5 \times 1.5 \times 1.5$ mm^3 [203].

- Bodurka et al. determine the optimal voxel size by matching the contribution of thermal noise to that of physiological noise as to optimise activation induced signal changes [204].

One of the challenges of fMRI is to explore the brain at submillimetre resolution. This would allow researchers to investigate the complex interactions between the

functional sub-units of the human brain. Hoogenraad et al. performed a study with submillimetre resolution and concluded that the activated areas were predominantly found to be in the sulci in the location of venous vessels with diameters on the order of the pixel size [205]. Attempts to use conventional fMRI techniques for high spatial resolution applications are hampered by technical limitations of the MR scanner and the spatial limitations of the HRs.

Mapping $CMRO_2$ changes

The fast response associated with the initial dip is potentially extremely important for BOLD response localisation. Malonek and Grinvald suggested this fast response is the result of a rapid increase in $CMRO_2$ before the flow has begun to increase, leading to an initial deoxyhaemoglobin increase [39]. If this interpretation is correct, the initial dip of the BOLD signal could provide a much more accurate map of neuronal activity in fMRI experiments. Indeed, this early $CMRO_2$ is better localised than the later positive BOLD signal, which includes contributions from draining veins. Results of a visual stimulus experiment using an EPI acquisition at 4 T by Menon et al. seemed to confirm this finding [52]. More recently, fMRI studies at high magnetic field strengths of 4.7 T and 9.4 T have demonstrated that the initial dip is detectable within cat visual cortex and can be used to resolve details on a spatial scale of 100-200 μm. [190, 206]. However, the initial dip is not always seen, and this has led to controversy over its existence [57, 58].

1.8 Conclusions

Since its introduction in the early 1990s, BOLD fMRI is widely used to identify the spatial layout of brain activation associated with sensory stimulations, motor actions and cognitive tasks. The technique tries to measure the neuronal activity indirectly, by means of the haemodynamic effect. The latter effect however, is associated with activation through a neurovascular coupling composed of a cascade of complex and poorly understood mechanisms. Moreover, the resulting signal change is small and exhibits a nonlinear component. The application of specific acquisition and processing techniques however, made BOLD fMRI a powerful approach to map brain functionality. Furthermore, the design of event-related paradigms overcame specific problems related to complex tasks.

Nevertheless, various confounding effects, both neuronal or non-neuronal in origin, as well as different noise sources contribute to the fMRI response. Several approaches exist to deal with these effects, such as an appropriate experimental design or correction with preprocessing techniques. However, a post-hoc interpretation of the detected activation patterns is mostly indispensable. Remaining artefactual components are as yet identified and subsequently excluded from further analysis.

The latter interpretation has to be performed with great care, as it introduces a subjective component in the analysis. Therefore, techniques for the analysis of fMRI data sets are assessed not only in terms of detection accuracy and robustness. Features such as the validability and interpretability of the results are as important. This combined approach will be the guideline for the assessment and comparison of techniques for the analysis of complex fMRI data sets. This topic is covered in chapter 2.

Bibliography

Bibliography

[1] C S Roy and C S Sherrington. On the regulation of the blood supply of the brain. *Journal of Physiology*, 11:85–108, 1890.

[2] S S Kety and C F Schmidt. The nitrous oxide method for the quantificative determination of cerebral blood flow in man: theory, procedure and normal values. *J Clin Invest*, 27(4):476–483, Jul 1948.

[3] W D Obrist et al. Determination of regional cerebral blood flow by inhalation of 133-Xenon. *Circ Res*, 20(1):124–135, Jan 1967.

[4] R S Frackowiak et al. Quantitative measurement of regional cerebral blood flow and oxygen metabolism in man using 15O and positron emission tomography: theory, procedure, and normal values. *J Comput Assist Tomogr*, 4(6):727–736, Dec 1980.

[5] P T Fox and M E Raichle. Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proc Natl Acad Sci USA*, 83(4):1140–1144, Feb 1986.

[6] F. Bloch et al. Nuclear induction. *Phys. Rev.*, 69(3-4):127, Feb 1946.

[7] E. M. Purcell et al. Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.*, 69(1-2):37–38, Jan 1946.

[8] P C Lauterbur. Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, 242:190–191, 1973.

[9] R.B. Buxton. *"Introduction to Functional Magnetic Resonance Imaging: principles and techniques"*. Cambridge University Press, 2001.

[10] J W Belliveau et al. Functional mapping of the human visual cortex by magnetic resonance imaging. *Science*, 254(5032):716–719, Nov 1991.

[11] S Ogawa et al. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci USA*, 87(24):9868–9872, Dec 1990b.

[12] K K Kwong et al. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc Natl Acad Sci USA*, 89 (12):5675–5679, Jun 1992.

[13] A Smolders et al. Dissecting cognitive stages with time-resolved fMRI data: a comparison of fuzzy clustering and independent component analysis. *Magn Reson Imaging*, May 2007.

[15] A Villringer and U Dirnagl. Coupling of brain activity and cerebral blood flow: basis of functional neuroimaging. *Cerebrovasc Brain Metab Rev*, 7(3): 240–276, Fall 1995.

[14] G M Boynton et al. Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci*, 16(13):4207–4221, Jul 1996.

[16] C Grimm et al. A comparison between electric source localisation and fMRI during somatosensory stimulation. *Electroencephalogr Clin Neurophysiol*, 106(1):22–29, Jan 1998.

[17] P S Churchland and T J Sejnowski. Perspectives on cognitive neuroscience. *Science*, 242(4879):741–5, 1988.

[18] K H Britten et al. A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis Neurosci*, 13(1):87–100, 1996.

[19] P Lennie. Single units and visual cortical organization. *Perception*, 27(8): 889–935, 1998.

[20] G M Boynton et al. Neuronal basis of contrast discrimination. *Vision Res*, 39(2):257–69, 1999.

[21] M Jueptner and C Weiller. Review: does measurement of regional cerebral blood flow reflect synaptic activity? Implications for PET and fMRI. *Neuroimage*, 2(2):148–56, 1995.

[22] M N Shadlen et al. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J Neurosci*, 16(4):1486–510, 1996.

[23] M L Platt and P W Glimcher. Neural correlates of decision variables in parietal cortex. *Nature*, 400(6741):233–8, 1999.

[24] A Bradley et al. Visual orientation and spatial frequency discrimination: a comparison of single neurons and behavior. *J Neurophysiol*, 57(3):755–72, 1987.

[25] J G Nicholls et al. *From Neuron to Brain*. Sinauer, Sunderland, MA, 1992.

[26] R L Grubb et al. The effect of changes in PCO2 on cerebral blood volume, blood flow, and vascular mean transit time. *Stroke*, 5:630–639, 1974.

[27] A Villringer et al. Dynamic imaging with lanthanide chelates in normal brain: contrast due to magnetic susceptibility effects. *Magn Reson Med*, 6 (2):164–174, Feb 1988.

[28] B R Rosen et al. Perfusion imaging by nuclear magnetic resonance. *Magn Reson Q*, 5(4):263–281, Oct 1989.

[29] J B Mandeville et al. Dynamic functional imaging of relative cerebral blood volume during rat forepaw stimulation. *Magn Reson Med*, 39(4):615–624, Apr 1998.

[30] K Scheffler et al. Titration of the BOLD effect: separation and quantitation of blood volume and oxygenation changes in the human cerebral cortex during neuronal activation and ferumoxide infusion. *Magn Reson Med*, 42(5): 829–36, 1999.

[31] L Pauling and C D Coryell. The magnetic properties and structure of Hemoglobin, Oxyhemoglobin and Carbonmonoxyhemoglobin. *Proc Natl Acad Sci USA*, 22(4):210–216, 1936.

[32] R M Weisskoff and S Kiihne. MRI susceptometry: image-based measurement of absolute susceptibility of MR contrast agents and human blood. *Magn Reson Med*, 24(2):375–383, Apr 1992.

[33] P T Fox and M E Raichle. Stimulus rate dependence of regional cerebral blood flow in human striate cortex, demonstrated by positron emission tomography. *Journal of Neurophysiology*, 51(5):1109–1120, 1984.

[34] R D Hoge et al. Investigation of BOLD signal dependence on cerebral blood flow and oxygen consumption: the deoxyhemoglobin dilution model. *Magn Reson Med*, 42(5):849–863, Nov 1999.

[35] R D Hoge et al. Linear coupling between cerebral blood flow and oxygen consumption in activated human cortex. *Proc Natl Acad Sci USA*, 96(16): 9403–9408, Aug 1999.

[36] R Turner et al. Echo-planar time course MRI of cat brain oxygenation changes. *Magn Reson Med*, 22(1):159–166, Nov 1991.

[37] P T Fox et al. Nonoxidative glucose consumption during focal physiologic neural activity. *Science*, 241(4864):462–464, Jul 1988.

[38] J W Prichard and B R Rosen. Functional study of the brain by NMR. *J Cereb Blood Flow Metab*, 14(3):365–372, May 1994.

[39] D Malonek and A Grinvald. Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: implications for functional brain mapping. *Science*, 272(5261):551–554, Apr 1996.

[40] R B Buxton and L R Frank. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *J Cereb Blood Flow Metab*, 17(1):64–72, Jan 1997.

[41] A Gjedde et al. Is oxygen diffusion limiting for blood-brain barrier transfer of oxygen? In N A Lassen et al., editors, *Brain Work and Mental Activity. Alfred Benzon Symposium*, pages 177–184, Copenhagen, 1991.

[42] I G Kassissia et al. Tracer oxygen distribution is barrier-limited in the cerebral microcirculation. *Circ Res*, 77(6):1201–1211, Dec 1995.

[43] D Bereczki et al. Hypoxia increases velocity of blood flow through parenchymal microvascular systems in rat brain. *J Cereb Blood Flow Metab*, 13(3): 475–486, May 1993.

[44] U Gobel et al. Lack of capillary recruitment in the brains of awake rats during hypercapnia. *J Cereb Blood Flow Metab*, 9(4):491–499, Aug 1989.

[45] J Frahm et al. Dynamic uncoupling and recoupling of perfusion and oxidative metabolism during focal brain activation in man. *Magn Reson Med*, 35(2): 143–148, Feb 1996.

[46] J B Mandeville et al. Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. *J Cereb Blood Flow Metab*, 19(6):679–689, Jun 1999.

[47] R B Buxton et al. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn Reson Med*, 39(6):855–864, Jun 1998.

[48] J B Mandeville et al. MRI measurement of the temporal evolution of relative $CMRO_2$ during rat forepaw stimulation. *Magn Reson Med*, 42(5):944–951, Nov 1999.

[49] P J Magistretti and L Pellerin. Cellular mechanisms of brain energy metabolism and their relevance to functional brain imaging. *Philos Trans R Soc Lond B Biol Sci*, 354(1387):1155–1163, Jul 1999.

[50] G Marchal et al. Regional cerebral oxygen consumption, blood flow, and blood volume in healthy human aging. *Arch Neurol*, 49(10):1013–1020, Oct 1992.

[51] T L Davis et al. Calibrated functional MRI: mapping the dynamics of oxidative metabolism. *Proc Natl Acad Sci USA*, 95(4):1834–1839, Feb 1998.

[52] R S Menon et al. BOLD based functional MRI at 4 Tesla includes a capillary bed contribution: echo-planar imaging correlates with previous optical imaging using intrinsic signals. *Magn Reson Med*, 33(3):453–459, Mar 1995.

[53] P A Bandettini. *MRI studies of brain activation: dynamic characteristics. In: functional MRI of the Brain*, pages 144–151. Society of Magnetic Resonance in Medicine, Berkeley, 1993.

[54] T Ernst and J Hennig. Observation of a fast response in functional MR. *Magn Reson Med*, 32(1):146–149, Jul 1994.

[55] X Hu et al. Evaluation of the early response in fMRI in individual subjects using short stimulus duration. *Magn Reson Med*, 37(6):877–884, Jun 1997.

[56] E Yacoub and X Hu. Detection of the early negative response in fMRI at 1.5 Tesla. *Magn Reson Med*, 41(6):1088–1092, Jun 1999.

[57] J J Marota et al. Investigation of the early response to rat forepaw stimulation. *Magn Reson Med*, 41(2):247–252, Feb 1999.

[58] P Fransson et al. Temporal characteristics of oxygenation-sensitive MRI responses to visual activation in humans. *Magn Reson Med*, 39(6):912–919, Jun 1998.

[59] K D Merboldt et al. Functional MRI of human brain activation combining high spatial and temporal resolution by a CINE FLASH technique. *Magn Reson Med*, 34(4):639–644, Oct 1995.

[60] G Kruger et al. Dynamic MRI sensitized to cerebral blood oxygenation and flow during sustained activation of human visual cortex. *Magn Reson Med*, 35(6):797–800, Jun 1996.

[61] A M Dale and R L Buckner. Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, 5:329–340, 1997.

[62] A L Vasquez and D C Noll. Nonlinear aspects of the BOLD response in functional MRI. *Neuroimage*, 7:108–118, 1998.

[63] G H Glover. Deconvolution of impulse response in event-related fMRI. *Neuroimage*, 9:416–429, 1999.

[64] M D Robson et al. Measurements of the temporal fMRI response of the human auditory cortex to trains of tones. *Neuroimage*, 7(3):185–198, Apr 1998.

[65] E D Adrian. The impulses produced by sensory nerve endings: Part I. *J Physiol*, 61(1):49–72, Mar 1926.

[66] A B Bonds. Temporal dynamics of contrast gain in single cells of the cat striate cortex. *Vis Neurosci*, 6(3):239–255, Mar 1991.

[67] T Maddess et al. Factors governing the adaptation of cells in area-17 of the cat visual cortex. *Biol Cybern*, 59(4-5):229–236, 1988.

[68] S A Engel et al. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex*, 7(2):181–92, 1997.

[69] S Ogawa et al. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magn Reson Med*, 14(1):68–78, Apr 1990a.

[70] S Ogawa and T M Lee. Magnetic resonance imaging of blood vessels at high fields: in vivo and in vitro measurements and image simulation. *Magn Reson Med*, 16(1):9–18, Oct 1990.

[71] P Mansfield. Multi-planar image formation using NMR spin echoes. *J. Phys. C:Solid State Phys.*, 10:L55–L58, 1977.

[72] J P J Pinel. *Biopsychology*. Pearson, 1999.

[73] J Talairach and P Tournoux. *Co-planar Stereotaxic Atlas of the Human Brain*. New York: Thieme Medical Publications, 1988.

[74] R M Hinke et al. Functional magnetic resonance imaging of Broca's area during internal speech. *Neuroreport*, 4(6):675–678, Jun 1993.

[75] J R Binder et al. Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci*, 17(1):353–62, 1997.

[76] A C Nobre et al. Word recognition in the human inferior temporal lobe. *Nature*, 372(6503):260–3, 1994.

[78] J Frahm et al. Dynamic MR imaging of human brain oxygenation during rest and photic stimulation. *J Magn Reson Imaging*, 2(5):501–505, Sep 1992.

[79] R S Menon et al. Functional brain mapping using magnetic resonance imaging. Signal changes accompanying visual stimulation. *Invest Radiol*, 27 Suppl 2:47–53, Dec 1992.

[80] A M Blamire et al. Dynamic mapping of the human visual cortex by high-speed magnetic resonance imaging. *Proc Natl Acad Sci USA*, 89(22):11069–11073, Nov 1992.

[77] S Ogawa et al. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci USA*, 89(13):5951–5955, Jul 1992.

[81] S Ogawa et al. Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. *Biophys J*, 64(3):803–812, Mar 1993b.

[82] R Turner et al. Functional mapping of the human visual cortex at 4 and 1.5 tesla using deoxygenation contrast EPI. *Magn Reson Med*, 29(2):277–279, Feb 1993.

[83] R T Constable et al. Functional brain imaging at 1.5 T using conventional gradient echo MR imaging techniques. *Magn Reson Imaging*, 11(4):451–459, 1993.

[84] A Connelly et al. Functional mapping of activated human primary cortex with a clinical MR imaging system. *Radiology*, 188(1):125–130, Jul 1993.

[85] W Schneider et al. Functional topographic mapping of the cortical ribbon in human vision with conventional MRI scanners. *Nature*, 365(6442):150–153, Sep 1993.

[86] C R Jr Jack et al. Sensory motor cortex: correlation of presurgical mapping with functional MR imaging and invasive cortical mapping. *Radiology*, 190 (1):85–92, Jan 1994.

[88] S G Kim et al. Functional magnetic resonance imaging of motor cortex: hemispheric asymmetry and handedness. *Science*, 261(5121):615–617, Jul 1993a.

[89] S G Kim et al. Functional imaging of human motor cortex at high magnetic field. *J Neurophysiol*, 69(1):297–302, Jan 1993b.

[90] S M Rao et al. Functional magnetic resonance imaging of complex human movements. *Neurology*, 43(11):2311–2318, Nov 1993.

[87] P A Bandettini et al. Time course EPI of human brain function during task activation. *Magn Reson Med*, 25(2):390–398, Jun 1992.

[91] E Formisano et al. Localisation and characterisation of auditory perception through Functional Magnetic Resonance Imaging. *Technol Health Care*, 6 (2-3):111–123, Sep 1998.

[92] F Di Salle et al. Functional fields in human auditory cortex revealed by time-resolved fMRI without interference of EPI noise. *Neuroimage*, 13(2): 328–338, Feb 2001.

[93] F Di Salle et al. fMRI of the auditory system: understanding the neural basis of auditory gestalt. *Magn Reson Imaging*, 21(10):1213–1224, Dec 2003.

[94] G McCarthy et al. Echo-planar magnetic resonance imaging studies of frontal cortex activation during word generation in humans. *Proc Natl Acad Sci USA*, 90(11):4952–4956, Jun 1993.

[95] S G Kim et al. Activation of a cerebellar output nucleus during cognitive processing. *Science*, 265(5174):949–951, Aug 1994.

[96] A Karni et al. Functional MRI evidence for adult motor cortex plasticity during motor skill learning. *Nature*, 377(6545):155–158, Sep 1995.

[97] B A Shaywitz et al. Sex differences in the functional organization of the brain for language. *Nature*, 373(6515):607–609, Feb 1995.

[98] M I Sereno et al. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, 268(5212):889–893, May 1995.

[99] E A DeYoe et al. Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc Natl Acad Sci USA*, 93(6):2382–2386, Mar 1996.

[100] R B Tootell et al. New images from human visual cortex. *Trends in Neuroscience*, 19:481–489, 1996.

[101] R B Tootell et al. From retinotopy to recognition: fMRI in human visual cortex. *Trends in cognitive science*, 2:174–183, 1998c.

[102] A Paivio. Comparisons of mental clocks. *J Exp Psychol Hum Percept Perform*, 4(1):61–71, Feb 1978. Comparative Study.

[103] R N Shepard and J Metzler. Mental rotation of three-dimensional objects. *Science*, 171(972):701–703, Feb 1971.

[105] M S Cohen et al. Changes in cortical activity during mental rotation. A mapping study using functional MRI. *Brain*, 119:89–100, 1996.

[106] M D'Esposito et al. A functional MRI study of mental image generation. *Neuropsychologia*, 35(5):725–730, May 1997.

[107] L Trojano et al. Matching two imagined clocks: the functional anatomy of spatial analysis in the absence of visual stimulation. *Cereb Cortex*, 10(5): 473–481, May 2000.

[104] E Formisano et al. Tracking the mind's image in the brain I: time-resolved fMRI during visuospatial mental imagery. *Neuron*, 35(1):185–194, Jul 2002.

[108] G Ganis et al. Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Brain Res Cogn Brain Res*, 20(2):226–241, Jul 2004.

[109] M A Just et al. Imagery in sentence comprehension: an fMRI study. *Neuroimage*, 21(1):112–124, 2004.

[110] A J King. Auditory neuroscience: activating the cortex without sound. *Curr Biol*, 16(11):410–411, Jun 2006.

[112] W Richter et al. Motor area activity during mental rotation studied by time-resolved single-trial fMRI. *J Cogn Neurosci*, 12(2):310–20, 2000.

[113] G Vingerhoets et al. Motor imagery in mental rotation: an fMRI study. *Neuroimage*, 17(3):1623–1633, Nov 2002.

[111] C Windischberger et al. Human motor cortex activity during mental rotation. *Neuroimage*, 20(1):225–232, Sep 2003.

[114] I Klein et al. Transient activity in the human calcarine cortex during visual-mental imagery: an event-related fMRI study. *J Cogn Neurosci*, 12 Suppl 2 (NIL):15–23, 2000.

[115] S M Kosslyn and W L Thompson. When is early visual cortex activated during visual mental imagery? *Psychol Bull*, 129(5):723–46, 2003.

[116] S D Slotnick et al. Visual mental imagery induces retinotopically organized activation of early visual areas. *Cereb Cortex*, 15(10):1570–1583, Oct 2005.

[117] P J Lang et al. Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3):261–273, May 1993.

[118] S M Kosslyn et al. Neural foundations of imagery. *Nat Rev Neurosci*, 2(9): 635–642, Sep 2001.

[119] J R Binder and S M Rao. *Human brain mapping with functional magnetic resonance imaging. In: Localization and Neuroimaging in Neuropsychology*, pages 185–212. Academic Press, San Diego, 1994.

[120] M D'Esposito et al. Event-related functional MRI: implications for cognitive psychology. *Psychol Bull*, 125(1):155–64, 1999.

[121] S Laurian et al. Short-term habituation of auditory evoked responses in schizophrenics. *Res. Commun. Psychol. Psychiatry Behav.*, 13:35–42, 1988.

[122] R L Buckner et al. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc Natl Acad Sci USA*, 93(25):14878–83, 1996.

[123] K J Friston et al. Stochastic designs in event-related fMRI. *Neuroimage*, 10 (5):607–19, 1999.

[124] R S Menon et al. Spatial and temporal resolution of functional magnetic resonance imaging. *Biochem Cell Biol*, 76(2-3):560–71, 1998b.

[125] F M Miezin et al. Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage*, 11(6):735–59, 2000.

[126] R S Menon et al. Mental chronometry using latency-resolved functional MRI. *Proc Natl Acad Sci USA*, 95(18):10902–10907, Sep 1998a.

[127] P A Bandettini. *The temporal resolution of MRI. In: functional MRI*, pages 205–220. Springer-Verlag, Mauer, Germany, 1999.

[128] R L Savoy et al. Pushing the temporal resolution of fMRI: Studies of very brief visual stimuli, onset variability and asynchrony, and stimulus-correlated change in noise. In *Proceedings of the Society of Magnetic Resonance Third Scientific Meeting and Exhibition*, volume 2, pages 450–..., 1995.

[129] P J Laurienti et al. Separating neural processes using mixed event-related and epoch-based fMRI paradigms. *J Neurosci Methods*, 131(1-2):41–50, Dec 2003.

[130] H Gudbjartsson and S Patz. The Rician distribution of noisy MRI data. *Magn Reson Med*, 34(6):910–914, Dec 1995.

[131] J Sijbers and A J den Dekker. Maximum likelihood estimation of signal amplitude and noise variance from MR data. *Magn Reson Med*, 51(3):586–594, Mar 2004.

[132] A J den Dekker and J Sijbers. Implications of the Rician distribution for fMRI generalized likelihood ratio tests. *Magn Reson Imaging*, 23(9):953–959, Nov 2005.

[133] J Sijbers et al. Automatic estimation of the noise variance from the histogram of a magnetic resonance image. *Phys Med Biol*, 52(5):1335–1348, Mar 2007.

[134] D I Hoult and R E Richards. The signal to noise ratio of the nuclear magnetic resonance experiment. *J Magn Reson*, 24:71–85, 1979.

[135] A Macovski. Noise in MRI. *Magn Reson Med*, 36(3):494–497, Sep 1996.

[136] D L Parker and G T Gullberg. Signal-to-noise efficiency in magnetic resonance imaging. *Med Phys*, 17(2):250–257, Mar 1990.

[137] P E Grant et al. High-resolution imaging of the brain. *Magn Reson Imaging Clin N Am*, 6(1):139–154, Feb 1998.

[138] P Jezzard et al. *Functional MRI. An introduction to Methods.* Oxford University Press, 2001.

[139] R E Hendrick et al. Optimizing tissue contrast in magnetic resonance imaging. *Magn Reson Imaging*, 2(3):193–204, 1984.

[140] F W Wehrli et al. The dependence of nuclear magnetic resonance (NMR) image contrast on intrinsic and pulse sequence timing parameters. *Magn Reson Imaging*, 2(1):3–16, 1984.

[141] K J Friston et al. To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. *Neuroimage*, 12(2):196–208, Aug 2000.

[142] O. Friman et al. Detection and detrending in functional MRI data analysis. *NeuroImage*, 22(2):645–655, 2004.

[143] X Descombes et al. fMRI signal restoration using a spatio-temporal Markov Random Field preserving transitions. *Neuroimage*, 8(4):340–349, Nov 1998.

[144] B A Ardekani et al. Activation detection in functional MRI using subspace modeling and maximum likelihood estimation. *IEEE Trans Med Imaging*, 18(2):101–114, Feb 1999.

[145] J Mayhew et al. Spectroscopic analysis of changes in remitted illumination: the response to increased neural activity in brain. *Neuroimage*, 10(3):304–326, Sep 1999.

[146] J V Hajnal et al. Artifacts due to stimulus correlated motion in functional imaging of the brain. *Magn Reson Med*, 31(3):283–291, Mar 1994.

[147] P L Purdon and R M Weisskoff. Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. *Hum Brain Mapp*, 6(4):239–249, 1998.

[148] E Zarahn et al. Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *Neuroimage*, 5 (3):179–197, Apr 1997.

[149] A M Smith et al. Investigation of low frequency drift in fMRI signal. *Neuroimage*, 9(5):526–533, May 1999.

[150] T R Oakes et al. Comparison of fMRI motion correction software tools. *Neuroimage*, 28(3):529–543, 2005.

[151] E Formisano et al. Cortex-based independent component analysis of fMRI time series. *Magn Reson Imaging*, 22(10):1493–1504, 2004.

[152] B Biswal et al. Reduction of physiological fluctuations in fMRI using digital filters. *Magn Reson Med*, 35(1):107–113, Jan 1996.

[153] L R Frank et al. Probabilistic analysis of functional magnetic resonance imaging data. *Magn Reson Med*, 39(1):132–148, Jan 1998.

[154] K Restom et al. Physiological noise reduction for arterial spin labeling functional MRI. *Neuroimage*, 31(3):1104–1115, Jul 2006.

[155] C Triantafyllou et al. Effect of spatial smoothing on physiological noise in high-resolution fMRI. *Neuroimage*, 32(2):551–557, Aug 2006.

[156] R B Buxton et al. Application of the balloon model to the bold response to stimuli of different duration. In *7th Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, pages 1735–.., Philadelphia, 1999.

[157] T L Davis et al. Susceptibility contrast undershoot is not matched by inflow contrast undershoot. In *SMR, 2nd Annual Meeting*, pages 435–.., San Fransisco, 1994.

[158] G L Shulman et al. Common blood flow changes across visual tasks: decreases in cerebral cortex. *J Cog Neurosci*, 9:648–663, 1997.

[159] M E Raichle et al. A default mode of brain function. *Proc. Natl. Acad. Sci. USA*, 98:676–682, 2001.

[160] A T Smith et al. Attentional suppression of activity in the human visual cortex. *Neuroreport*, 11(2):271–277, 2000.

[161] Z S Saad et al. Analysis and use of fMRI response delays. *Human Brain Mapping*, 13(2):74–93, 2001.

[162] R B Tootell et al. The retinopy of visual spatial attention. *Neuron*, 21(6): 1409–1422, 1998a.

[163] R B Tootell et al. The representation of the ipsilateral visual field in human cerebral cortex. In *Proc Natl Acad Sci USA, Feb(3)*, volume 95, pages 818–824, 1998b.

[164] P H Ghatan et al. Coexistence of attention-based facilitation and inhibition in the human cortex. *Neuroimage*, 7(1):23–29, 1998.

[165] M Hutchinson et al. Task-specific deactivation patterns in functional magnetic resonance imaging. *Magn Reson Imaging*, 17(10):1427–1436, 1999.

[166] M E Raichle. Behind the scenes of functional brain imaging: a historical and physiological perspective. In *Proc Natl Acad Sci USA, Feb(3)*, volume 95, pages 765–772, 1998.

[167] A Shmuel et al. Sustained negative BOLD, blood flow and oxygen consumption response and its coupling to the positive response in the human brain. *Neuron*, 36(6):1195–1210, 2002.

[168] N Harel et al. Origin of negative blood oxygenation level-dependent fMRI signals. 22(8):908–917, 2002.

[169] A T Smith et al. Negative BOLD in the visual cortex: evidence against blood stealing. *Human Brain Mapping*, 21(4):213–220, 2004.

[170] G Gomiscek et al. A possible role of in-flow effects in functinal mr-imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 1(3-4): 109–113, 1993.

[171] J H Duyn et al. Inflow versus deoxyhemoglobin effects in BOLD functional MRI using gradient echoes at 1.5 T. *NMR Biomed*, 7(1-2):83–8, 1994.

[172] J Frahm et al. Brain or vein–oxygenation or flow? On signal physiology in functional MRI of human brain activation. *NMR Biomed*, 7(1-2):45–53, 1994.

[173] D Le Bihan et al. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology*, 168(2):497–505, Aug 1988.

[174] B LeBihan. *Diffusion and perfusion magnetic resonance imaging*. Raven Press, New York, 1995.

[175] J L Boxerman et al. The intravascular contribution to fMRI signal change: Monte Carlo modeling and diffusion-weighted studies in vivo. *Magn Reson Med*, 34(1):4–10, Jul 1995.

[176] D J Dubowitz et al. Functional magnetic resonance imaging in macaque cortex. *Neuroreport*, 9(10):2213–2218, Jul 1998.

[177] L Stefanacci et al. fMRI of monkey visual cortex. *Neuron*, 20(6):1051–1057, Jun 1998.

[178] N K Logothetis et al. Functional imaging of the monkey brain. *Nat Neurosci*, 2(6):555–562, Jun 1999.

[179] J G Ojemann et al. Functional MRI studies of word-stem completion: reliability across laboratories and comparison to blood flow imaging with PET. *Hum Brain Mapp*, 6(4):203–15, 1998.

[180] V P Clark et al. Functional magnetic resonance imaging of human visual cortex during face matching: a comparison with positron emission tomography. *Neuroimage*, 4(1):1–15, 1996.

[181] B J Casey et al. Reproducibility of fMRI results across four institutions using a spatial working memory task. *Neuroimage*, 8(3):249–61, 1998.

[182] Nikos K Logothetis and Brian A Wandell. Interpreting the BOLD signal. *Annu Rev Physiol*, 66:735–769, 2004.

[183] G K Aguirre et al. The variability of human, BOLD hemodynamic responses. *Neuroimage*, 8(4):360–9, 1998.

[184] R L Buckner et al. Functional-anatomic study of episodic retrieval. II. Selective averaging of event-related fMRI trials to test the retrieval success hypothesis. *Neuroimage*, 7(3):163–75, 1998.

[185] D L Schacter et al. Late onset of anterior prefrontal activity during true and false recognition: an event-related fMRI study. *Neuroimage*, 6(4):259–69, 1997.

[186] K R Thulborn et al. Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field. *Biochim Biophys Acta*, 714(2):265–270, Feb 1982.

[187] S Lai et al. Identification of vascular structures as a major source of signal contrast in high resolution 2D and 3D functional activation imaging of the motor cortex at 1.5T: preliminary results. *Magn Reson Med*, 30(3):387–392, Sep 1993.

[188] D C Alsop and J A Detre. Reduced transit-time sensitivity in noninvasive magnetic resonance imaging of human cerebral blood flow. *J Cereb Blood Flow Metab*, 16(6):1236–1249, Nov 1996.

[189] E C Wong et al. Implementation of quantitative perfusion imaging techniques for functional brain mapping using pulsed arterial spin labeling. *NMR Biomed*, 10(4-5):237–249, Jun 1997.

[190] D S Kim et al. High-resolution mapping of iso-orientation columns by fMRI. *Nat Neurosci*, 3(2):164–9, 2000.

[191] J S Gati et al. Experimental determination of the BOLD field strength dependence in vessels and tissue. *Magn Reson Med*, 38(2):296–302, 1997.

[193] M S Cohen and S Y Bookheimer. Localization of brain function using magnetic resonance imaging. *Trends Neurosci*, 17(7):268–77, 1994.

[192] K K Kwong. Functional magnetic resonance imaging with echo planar imaging. *Magn Reson Q*, 11(1):1–20, Mar 1995.

[194] F Farzaneh et al. Analysis of T2 limitations and off-resonance effects on spatial resolution and artifacts in echo-planar imaging. *Magn Reson Med*, 14(1):123–39, 1990.

[195] C M Collins et al. SAR and B_1 field distributions in a heterogeneous human head model within a birdcage coil. specific energy absorption rate. *Magnetic Resonance in Medicine*, 40:847–856, 1998.

[196] F G Shellock et al. Determination of gradient magnetic field-induced acoustic noise associated with the use of echo planar and three-dimensional, fast spin echo techniques. *J Magn Reson Imaging*, 8(5):1154–1157, Sep 1998.

[197] B P Poncelet et al. Brain parenchyma motion: measurement with cine echo-planar MR imaging. *Radiology*, 185(3):645–51, 1992.

[198] G Kruger et al. Neuroimaging at 1.5 T and 3.0 T: comparison of oxygenation-sensitive magnetic resonance imaging. *Magn Reson Med*, 45(4):595–604, 2001.

[199] J M Oja et al. Venous blood effects in spin-echo fMRI of human brain. *Magn Reson Med*, 42(4):617–626, Oct 1999.

[200] P C van Zijl et al. Quantitative assessment of blood flow, blood volume and blood oxygenation effects in functional magnetic resonance imaging. *Nat Med*, 4(2):159–167, Feb 1998.

[201] S P Lee et al. Diffusion-weighted spin-echo fMRI at 9.4 T: microvascular/tissue contribution to BOLD signal changes. *Magn Reson Med*, 42(5): 919–928, Nov 1999.

[202] J Frahm et al. Functional MRI of human brain activation at high spatial resolution. *Magn Reson Med*, 29(1):139–144, Jan 1993.

[203] J S Hyde et al. Optimal voxel size in fMRI. In *Proceedings ISMRM 8th Annual Meeting*, pages 240–..., Denver, 2000.

[204] J Bodurka et al. Mapping the MRI voxel volume in which thermal noise matches physiological noise–implications for fMRI. *Neuroimage*, 34(2):542–549, Jan 2007.

[205] F G Hoogenraad et al. Sub-millimeter fMRI at 1.5 Tesla: correlation of high resolution with low resolution measurements. *J Magn Reson Imaging*, 9(3): 475–482, Mar 1999.

[206] T Q Duong et al. Spatiotemporal dynamics of the BOLD fMRI signals: toward mapping submillimeter cortical columns using the early negative response. *Magn Reson Med*, 44(2):231–42, 2000.

Chapter 2

Data-driven analysis techniques for fMRI data sets

2.1 Introduction

This chapter elaborately reviews and compares the applied techniques to analyse the acquired BOLD fMRI data. It elucidates the shift from a hypothesis-driven approach to a data-driven approach and focuses on two data-driven techniques currently favoured for the analysis of complex data sets: the Fuzzy Clustering Method and spatial Independent Component Analysis. This chapter illustrates their principles and clarifies their strengths and shortcomings and covers a literature review of the major applications in a historical perspective. Finally, the Receiver Operating Characteristics (ROC) technique to assess the inherent accuracy of a detection procedure is elucidated.

2.2 Hypothesis versus data-driven approach

Determining the precise location and time course of the responses corresponding to true neuronal activity is the subject of many BOLD fMRI studies. The acquired fMRI signals have no simple quantitative physiological interpretation but contain temporal and spatial structure. Basically two different signal processing strategies exist to extract this structure: hypothesis (or model)-driven and data-driven methods.

Hypothesis-driven techniques specify a priori a model of the haemodynamic response and its goodness-of-fit at each voxel is tested by statistical methods.

This approach is also called confirmatory analysis.

Data-driven techniques attempt to find common features within the data, without (strong) a priori assumptions about the time profile of the haemodynamic response. This approach is also called exploratory analysis.

Combining the strength of the data-driven and the hypothesis-driven techniques in a hybrid method may lead to a more powerful approach. We will now elaborate into detail on methods implementing these approaches.

2.2.1 Hypothesis-driven approach

Subtraction and t-test

In the first fMRI studies, a simple method based on image subtraction was used to create descriptive images of the task-dependent brain areas [1–3]. Voxels with high signal amplitude in the difference image, formed by subtracting the 'control' from the 'task' condition images, reflect the areas with a task induced differential activation. To remove artefacts due to voxels with highly variable signals, each measured difference signal is normalized by dividing by an estimate of the intrinsic variability of the signal from that voxel (see 1.6.1). Voxels with a large signal difference due only to the fact that they have a large intrinsic variance will be suppressed, whereas true activations in which the signal change is much larger than the intrinsic variance will remain. The final result is a statistical parametric map (SPM), i.e. a map of the SNR of the difference measurement [4, 5].

A standard statistical test used to quantify the quality of an activation is the t-statistic [2]. The Student t-statistic is a *parametric* statistic to test whether a significant difference can be found between two states, in this case rest and activation. In case of non-parametric data, a Kolmogorov-Smirnov test is performed [6]. For a complete and detailed description of statistical tests we refer to [7].

In the simplest form of the t-test, the signals measured from a particular voxel are treated as samples of two populations, namely 'active' and 'rest', *as defined by the stimulus*. The t-test assesses whether there is a significant difference between the means of the measured BOLD values, corresponding to these two groups. It is a measure of how large the difference of the means is, compared to the variability of the populations. The test is performed under the null hypothesis that the distribution of voxel values during the behavioural task is identical to that during performance of the experimental task. Subsequently, a significance threshold is selected, such as $p < 0.05$, expressing that the possibility that a measured effect will arise by chance alone is less than 5 %. This significance threshold corresponds to a threshold on t and the voxels which pass the test are labelled 'activated'. It is also worth mentioning that the value of the statistic itself should not be taken as a measure of the degree of activation: a larger value of the statistic does not imply a larger level of activation. To draw such a conclusion, one must show that the

difference in the statistic is due to a difference in the signal change, rather than a difference in the intrinsic variability of the signal.

Correlation analysis

The use of the stimulus function as a reference to detect HRs corresponding to activation is not optimal however. The BOLD response of a particular brain region does not precisely match the stimulus. BOLD fMRI provides a measure of the local, temporal pattern of neural activity, but only after that pattern has passed through a haemodynamic filter that smoothes and delays the signal. Bandettini et al. were the first to produce a reference time course by averaging the haemodynamic responses in a region of interest for a motor-task experiment [8]. They accounted for the shape and the delay of the response (Fig. 2.1). Each measured voxel time course was analysed by calculating the correlation coefficient between the data (after removal of linear drift) and the reference function. The correlation coefficient expresses the degree to which the measured signal follows the reference function. It is used as a statistical parameter for mapping once a threshold is selected (corresponding to a significance level) to determine activated voxels. When the MR signal from a voxel drifts (usually slightly) over the course of an experiment, simple linear detrending will enhance the accuracy of the correlation analysis. However, the time courses of processes related to changes in arousal, task strategy, head position, machine artefacts or other endogenous processes occurring during a trial may not resemble simple linear functions.

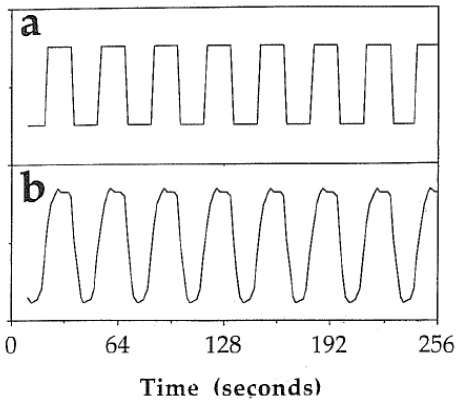

Fig. 2.1: *Stimulus paradigm (a) and correlation reference function (b) used in the study performed by Bandettini et al. [8].*

There is a one-to-one correspondence between the correlation coefficient and the t-test value. The simple t-test in which the time course data are divided into two groups and the means compared is equivalent to using a model reference function that precisely matches the stimulus pattern for correlation analysis. As indicated

before, the correlation analysis is more general, because a model function, reflecting the true response, is used instead of the square-wave stimulus pattern.

Modelling the HR function

Bandettini et al. produced a reference HR to detect motor activity by averaging the haemodynamic responses in a particular region of interest. However, this approach is difficult to generalise due to the variability and nonlinearity of the BOLD response, as discussed more elaborately in section 1.4. It was illustrated that the BOLD response exhibits a large variability between subjects [9] and even within subjects. The latter variability exists between multiple scans (the same day of different days spread out over a longer period, such as several months) and between cortical regions [9]. Moreover, the BOLD response is not a linear function of the stimulus duration [10–14].

Several authors proposed approaches to model the HR function *a priori*, as to accurately predict the shape of the fMRI signal (Fig. 2.2). Boynton et al. used a gamma function with 2 free parameters to model empirically derived HRs [11]. Friston et al. suggested to model the shape of the HR with a Poisson function [15]. Aguirre et al. proposed a model based on the first eigenvector of the HR in a study of the variability in the haemodynamic response within and between subjects [16]. Although the parameters of the proposed models are usually reasonably well known, it is worth estimating these parameters from the data [17–21].

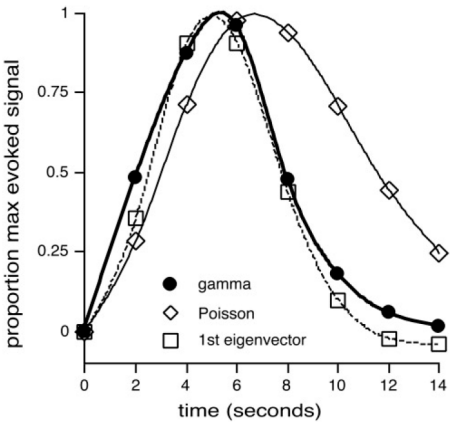

Fig. 2.2: *Three different models of the HR function. The thick solid line (circles) is described by a gamma function [11]. The thin solid line (diamonds) is based on a Poisson function [15]. The dotted line (squares) is base on the first eigenvector [16].*

General linear model

Correlation analysis applies only *one* model function. However, the analysis can be generalized to include any number of model functions each testing a factor of interest that may contribute to the fMRI signal in addition to the task design. Drift is also taken into account by including additional model functions which vary linearly or quadratic (or even following higher orders) in time. This basic multiple regression approach is referred to as the general linear model [15, 22, 23]. It is worth emphasizing that 'linear' refers to the fact that the data are modelled as a linear combination of model functions but that the model functions themselves need not be linear.

The GLM procedure determines the relative contribution of the modelled effects for each voxel and those voxels whose contribution to an effect exceeds a preselected threshold are labelled 'activated'. The residuals, i.e. the signals where the reference functions can't account for, are assumed to be independent and identically normally distributed with zero mean and constant variance. It can be shown that the t-test and correlation analysis (as well as most other parametric tests) can be regarded as special cases of the GLM.

2.2.2 Shortcomings of hypothesis-driven approach

A priori construction of the model

A hypothesis-driven approach is a powerful technique, but the accurate determination of a model of the expected HRs a priori can be difficult to accomplish. The influence of failure to accurately model the shape of the HR on sensitivity depends primarily on the design. In case of traditional block designs, in which relatively long duration periods of neural activity are evoked, the latter failure will likely result in only small decrements of sensitivity. For event-related fMRI designs, accurate estimation of the HR becomes more important to maintain sensitivity. For designs in which different trial types (task variance) are randomly ordered, the latter failure can be expected to result in substantial decrements in sensitivity, especially as the spacing between the trials decreases [10, 24].

Additionally, data sets originating from a complex task contain unexpected phenomena that are not modelled or phenomena that cannot be modelled a priori. Detection and characterisation of these phenomena is not possible with this technique.

Univariateness of the model

A hypothesis-driven approach mostly is univariate, i.e. voxel-based. They ignore interactions between voxels and are therefore expected to result in many false positives. Indeed, statistical tests are performed separately for a large number of time courses (on the order of 100,000 for whole brain coverage). Therefore, a large

amount of voxel comparisons will be statistically significant due to a change that has nothing to do with the task (for $p < 0.05$, on average 5,000 for whole brain coverage).

Several approaches exist to deal with this shortcoming.

Bonferroni correction The Bonferroni correction decreases the statistical threshold in proportion to the number of comparisons made [25]. It is however overly conservative, since it does not take into account for correlations among adjacent voxels, which are not statistically independent, but tend to be functionally similar.

False discovery rate Benjamini and Hochberg proposed a similar approach based on the 'false discovery rate' (FDR) [26, 27]. The latter technique is better suited for multiple comparisons, since it controls for the expected proportion of false positives across only those values reaching statistical significance, rather than controlling false positives for all voxels in the analysis that potentially could be significant due to chance.

Multivariate approach An alternative approach is the application of a multivariate technique, which takes advantage of the functional organization of the brain [28, 29]. Indeed, the brain processes information in a massively parallel network of highly interconnected neuronal ensembles. The latter is reflected by two complementary principles: localization and connectionism [30]. Localization implies that each psychomotor function is performed principally in a small set of brain areas. Connectionism posits that the brain regions involved in a given psychomotor function may be widely distributed. The brain activity required to perform a given task may thus be the functional integration of activity in multiple macroscopic loci or distinct brain systems.

The principle of localization derives originally from clinical experience where a restricted locus of damage to the nervous system could usually be inferred from a specific pattern of deficits demonstrated by a subject. The principle of connectionism was demonstrated in several studies determining the locus of a lesion. Occasionally, this could not accurately be accomplished, because the lesion interrupts connections between macroscopic loci required to perform some psychomotor task [31]. The principles of localisation and connectionism of the neural processes imply that the different physiological phenomena of interest as well as artefacts, may concern measurements in different brain regions.

In general, hypothesis-driven methods implement a univariate approach, but occasionally, multivariate extensions of existing univariate hypothesis-driven techniques were proposed. Friston et al. used standard multivariate statistics and the general linear model to make inferences about effects of interest when analysing fMRI data [4]. Canonical correlation analysis (CCA) is a

multivariate generalization of correlation analysis, measuring the linear relationship between multidimensional variables [32, 33].

2.2.3 Data-driven approach

A model-free and multivariate approach

Since hypothesis-driven approaches for the analysis of complex fMRI data sets often failed to extract the underlying neuronal responses, data-driven approaches were introduced. They do not require the a priori construction of a model of the underlying neuronal responses, but attempt to find common features within the data, without (strong) a priori assumptions about the data content itself. Moreover, they take advantage of the multivariate nature of the fMRI data set [34]. However, as their approach is exploratory instead of confirmatory, the reproducibility inter- and intra-subjects is a key issue.

Classes of data-driven methods

Most data-driven methods applied for the analysis of fMRI data sets can be divided into two categories, based on the approach used to elucidate the structure in the acquired data sets.

(1) A first class of methods transforms the original data as to separate signal from noise or to distinguish between different signals. At the same time, a reduction of the dimensionality is often obtained. Typical techniques implementing this approach are principal components analysis (PCA), factor analysis (FA), independent component analysis (ICA), wavelets [35–37], and autocorrelation maximisation [38].

(2) A second class of methods is based on a clustering approach. Clustering techniques classify the observed signals into several patterns according to similarity between each other. An overview of clustering methods can be found in [39]. Most clustering methods follow a partitioning-based approach, i.e. the data will be partitioned into an a priori determined number of clusters, thereby optimising an objective function or likelihood. The clusters that are thus identified may be exclusive, i.e. every element belongs to one group only (C-means). Or, they may be overlapping, meaning that an element may belong to several clusters to a certain degree (Fuzzy C-means). Or, they may be probabilistic, meaning that an element has a certain probability to belong to a cluster (Gaussian Mixture Model). As opposed to partitioning-based methods, techniques exist that do not need the number of clusters as preliminary information, such as hierarchical clustering techniques. They iteratively merge clusters that are the most similar into a larger structure. Alternatively, a combined split-merge approach can be applied.

The latter subdivision of analysis techniques is not exhaustive, however. Several other techniques exist, such as deterministic annealing [40], neural gas [41, 42], temporal clustering analysis [43], region-growing methods [44], and a self organising map (SOM) [45, 46].
It is worth mentioning that most methods are based on the *magnitude* data. Relatively few studies exist that are based on the *complex* data to take advantage of the phase information [47–49].

In this dissertation, the focus lies on FCM and ICA, two data-driven techniques for the analysis of magnitude fMRI data, originating from a complex fMRI experiment. In the following sections, we will elaborate on the principles of both methods, as well as on PCA, which is typically used as a preprocessing technique. Additionally, an application-oriented overview of literature is presented.

2.3 Preprocessing the data with Principal Components Analysis

2.3.1 Partitioning variance

One of the first data-driven approaches used to extract the relevant information out of an fMRI signal consisted in a transformation of the original data as to reduce the variance. This idea was initially implemented using the principal component analysis (PCA) technique, which explores and decomposes the (temporal or spatial) correlations present in the data set. The technique was first described in 1901 by Pearson [50] and developed independently by Hotelling [51]. It has since then been used widely for data decomposition, dimension reduction in signal, and image data compression and coding applications. In the latter domain it is also known as the Karhunen-Loève transform.

PCA was first applied in functional neuroimaging studies to PET data in a letter repetition and word generation task [52]. In the latter study, PCA revealed a component which could not (easily) be detected by a hypothesis-driven approach. Subsequently, experiments were performed in the field of fMRI [53–56].

In the analysis of fMRI data sets, PCA typically applies a spatial view on the data (see section 3.2). Spatial PCA tries to describe the acquired images as a linear combination of orthonormal spatial patterns, the so-called eigenimages. The first eigenimage represents the largest source of variance; the second represents the largest source of residual variance orthogonal to the first eigenimage and so on. This procedure is repeated up to a maximum number, determined by the rank of the data matrix. The images can be seen as maps of functional connectivity, because they share the same time pattern. The associated time patterns can be seen as modulation functions of the spatial components. PCA relegates most of the random noise to the trailing components, while the systematic structure is collected into the leading ones. Reducing the variance can thus be used to separate distinct responses and noise sources from each other. PCA also provides a useful

way to reduce the data dimensionality, because it allows the data to be represented by a limited number of principal components (PCs) to a specified level of accuracy (see section 3.3.1). Low-percentage signal variations, having a small influence on overall signal variance, can thereby be discarded.

2.3.2 Strength and shortcomings of PCA

PCA components are linear combinations of the original images and therefore usually do not correspond to real physical quantities, i.e. the images of the area activated by physiological stimulation. These physiologically relevant images, the so-called factors, are searched for in factor analysis (FA) by rotation of the PCA components. This is the reason why factor analysis is often called 'rotated principal component analysis'. The transformation of the factors need not to be orthogonal, but can also be oblique. An oblique transformation overcomes the limitation given by the orthogonality of PCA and separates the weak and mutually correlated (i.e. spatially overlapping) activated areas. This operation is a search amongst all linear combinations of the PCs and can be seen as a projection pursuit task. The interesting projections of the data can be found by using some a priori information about the expected result [Andersen 1999; Backfrieder 1996].

Both methods were occasionally successful but failed just as often in detecting task-related activations. The main reason for failure is that both PCA and FA segregate the data by partitioning its *total* variance into (un)correlated components. Performance-related fMRI changes, however, are often only a small part of the total signal variance. The influence of small activation regions may result in small percentages of the total variance, compared to other sources, both physiological and artefactual, distributed across brain voxels. Movement-related components for example can cause large signal variations in the data set, contributing heavily to the overall variance. Therefore, variance partitioning, whether using orthogonal or oblique components, does not always separate the data unambiguously into activated, noisy, and artefact-related components.

At present, PCA is mainly used as a preprocessing step to separate signal from noise and to reduce the number of dimensions. The relevance of the PCs is usually quantified by the percentage of variance explained by the components. Because usually the first PCs account for most of the variance in the signal, a threshold is set to capture most of the variance in the data (e.g. 90%) while removing the variance due to the noise. The latter threshold determines the number of components to be retained for further analysis. A larger percentage of the total data variance is now represented in a subspace of reduced dimension spanned by the dominant principal components.

2.4 Principles and applications of sICA for the analysis of fMRI data sets

2.4.1 Searching for statistical independence

Independent component analysis (ICA) is conceptually similar to PCA. Like PCA, ICA is a technique that models the data as a linear combination of components, expressing the underlying neural activity corresponding to a task. The main difference between the methods lies in the fact that in case of ICA the components are assumed to be statistically independent. Therefore, ICA also takes into account higher order statistics, whereas PCA is based on second order correlations only. Statistical independence proved to be useful as a guiding principle in a broad range of applications: telecommunication, image feature extraction, financial time series analysis, to artefacts separation in brain imaging [57].

2.4.1.1 History of ICA

In the early 1980s, techniques lying at the base of ICA were introduced by Hérault, Jutten, and Ans [58, 59], when trying to recover underlying signals from linear mixtures of those signals. The problem of unmixing these measured signals was known as blind source separation (BSS). Both the sources and the mixing process are unknown and the term 'blind' implies that such methods can separate data into source signals even if very little is known about the nature of those source signals. The first solutions to the BSS problem were presented by Comon [60] and Cardoso [61] during an international workshop on higher-order spectral analysis. Cardoso used higher-order cumulant tensors. A cumulant of an order is expressed as a combination of moments but has favourable properties. Among other advantages, a cumulant presents in a clearer way the additional information provided by higher order statistics. Several alternative approaches were proposed in the early 90s, based on nonlinear decorrelation [62], maximum likelihood [63, 64], or fourth-order cumulants [65].

In 1995, Bell and Sejnowski presented an information theoretic approach to extract the statistically independent components (ICs) [66]: 'InfoMax'. This algorithm maximizes the information in the components, which is equivalent to minimizing the mutual information between the components. The method was first applied to solve problems similar to the 'cocktail party' problem in which many people are speaking at once. The algorithm, assuming temporal independence (independence of the voices), could separate mixed signals into individual sources (voices). InfoMax proved to be a powerful new approach for the analysis of neuroimaging data sets. It was applied to the analysis of EEG data in which a set of signals, one from each electrode, are separated into temporally independent groups [67]. McKeown et al. were the first to analyse an fMRI data set with InfoMax [68]. They found that each of the separate processes (CTR, TTR, as well as confounding modes representing e.g. head motion) could be represented by one or more

spatially-independent components, each associated with a single time course and a spatial map. ICA subsequently broke through as a technique for the analysis of more complex fMRI data, holding several patterns of activation.

In 1997, Hyvärinen and Oja presented an information theoretic approach, based on the maximisation of negentropy, a concept based on entropy [69]. Their algorithm, 'FastICA' has contributed to the application of ICA to large-scale problems due to its computational efficiency based on the use of a fixed-point algorithm. The technique produced similar results as InfoMax [70].

2.4.1.2 Temporal and spatial ICA

As with PCA, ICA of fMRI data can be carried out in the temporal or spatial domain. In case of spatial ICA (sICA), the data set is decomposed into a set of spatial patterns of activations with associated time courses, assuming statistical independence among the spatial patterns. In case of temporal ICA (tICA), the data set is decomposed into a set of temporal patterns of activations with associated spatial maps, assuming statistical independence among the temporal patterns. Fig. 2.3 visually depicts both approaches for the analysis of fMRI data [71]. Data matrix $\boldsymbol{X}$, containing results of an fMRI experiment (i.e. the number of voxels N is much larger than the number of time samples T) is transposed when switched to the other approach. The yellow rectangles indicate the location of temporal information of a particular voxel in both views.

Generally, spatial ICA is chosen for the analysis of fMRI data sets. The latter choice involves two aspects.

- The choice depends on the agreement with both hypotheses of statistical independence. From a neuroscientific point of view, spatial independence is generally more plausible for fMRI data and temporal independence for EEG data. The experiment of McKeown et al., applying ICA for the first time to fMRI data instead of EEG data, introduced an abrupt change from tICA to sICA [68, 72].

- As an additional advantage, sICA is computationally less demanding, as independence has to be accomplished between a *limited* number of source images (i.e. the number of acquired time samples per experimental run). A data matrix resulting from an EEG experiment has a much larger temporal than spatial dimension. Functional MRI on the other hand results in a data set with much larger spatial than temporal dimension: typically 100.000 voxels for the whole brain, each holding a time course of hundreds of time points.

Although spatial ICA has by far dominated the functional imaging literature to date, some reports of the application of temporal ICA to fMRI data have appeared.

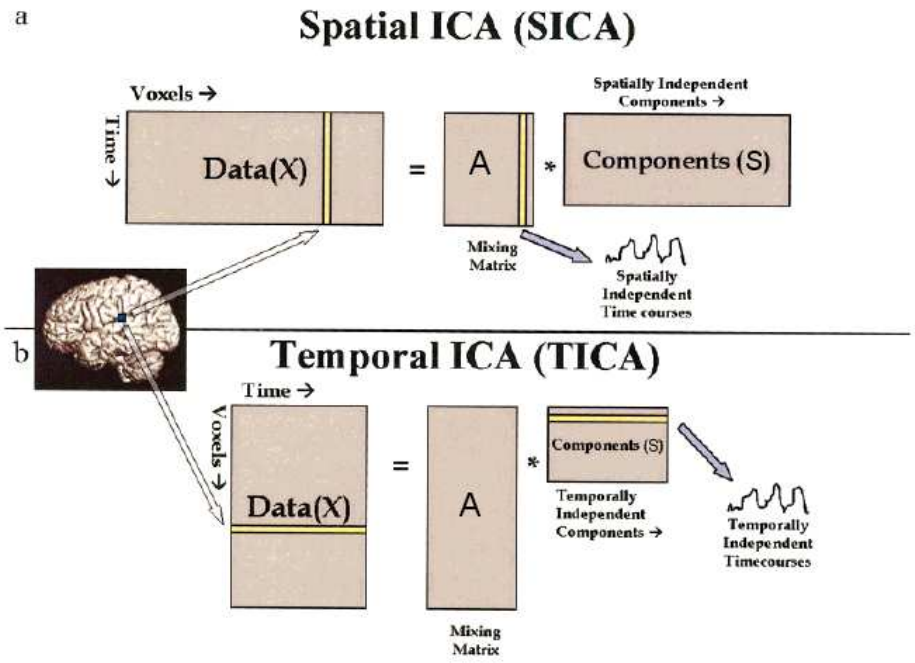

Fig. 2.3: *Visual depiction of the matrix representation of sICA (a) and tICA (b) approaches [71]. In each case, the spatial information is rearranged in one dimension (next to the temporal dimension). In sICA, the algorithm attempts to find spatially independent maps with associated time courses. In tICA, the algorithm attempts to find temporally independent time courses with associated spatial maps.*

Biswal and Ulmer found that tICA analysis was able to resolve two different induced effects on the fMRI signals, a task induced effect and CO_2 inhalation [73]. The use of tICA was motivated by the fact that the latter is a global effect, which makes sICA less suited. Peterson et al. suggested that sICA and tICA yield similar results for experiments in which there is one predictable task-related component [74]. However, the study also shows that sICA and tICA diverge if the predictable components are highly correlated in space or time, respectively. Calhoun et al. showed that both methods give similar results in decomposing data containing a pair of task-related waveforms, that are both spatially and temporally independent [71]. However, if the components are correlated in time, tICA will not separate all components and if the components are spatially correlated, sICA will not separate all components. In the remainder of this chapter, dealing with the analysis of fMRI data sets, spatial independence between the source maps will be pursued.

2.4.2 Strength and shortcomings of ICA

2.4.2.1 Comparison to hypothesis-driven techniques

ICA was compared to hypothesis-driven techniques on simulated and real fMRI data. For a comparison between ICA and FCM, we refer to section 2.5.2.1. A study performed by Lange et al. illustrated that ICA was able to identify features in brain maps not accessible by simple correlation or t-test based methods [75]. Other studies showed that ICA was also able to retrieve more complex patterns of brain activation often left undetected by PCA. More generally, ICA has proven to be remarkably versatile in several applications in which the brain activation was hard to predict beforehand. Activity in the visual [76–79], auditory [80], and cognitive domains [81], and even complex social interaction while simultaneously scanning more than one subject [82] have all been investigated with ICA.

Much work was done to assess the ability of ICA to reduce the contribution of noise sources to the fMRI signal. To that purpose, Thomas et al. explored and compared PCA and ICA-based techniques for the isolation of structured and random noise components [83]. After removal of these components, the images were reconstructed from the decomposed data sets. ICA proved better for the isolation and removal of structured noise, while PCA was superior for random noise. Perlbarg et al. corrected for structured noise in fMRI data by automatically identification of the ICA components [84]. Therefore, they used prior information on the spatial localisation of the main physiological fluctuations. A more detailed review can be found in [85].

2.4.2.2 Selection of parameters

Independent component analysis is an information theoretic approach yielding several algorithmic implementations, each characterised by specific parameters. Therefore, the elaboration on these parameters and the approach to determine a well-considered setting is performed in chapter 3 (see section 3.3.3, dealing with the implementing ICA algorithms). The determination the optimal 'number of ICs' is a common problem for all algorithms and is elucidated in section 2.4.2.3 in the context of the validation of the ICs.

2.4.2.3 Interpretation of the results

The interpretation of the results concerns the determination of maps of significant activation and their corresponding time courses out of the ICA results. Preliminary, however, the validation of the results, i.e. the selection of physiologically interesting ICs has to be accomplished. Both aspects are elucidated in this section. However, the interpretation may be hampered by the limited applicability of the ICA model in the analysis of a particular data set. The latter is therefore expounded as well.

Applicability of the model

Although ICA is a powerful technique for the analysis of fMRI data, it imposes some constraints: statistical independence is often only satisfied approximately, ICA offers a linear model [86], and it attempts to characterize the data globally. The latter means that even if the dataset is spatially heterogeneous, the ICA model tries to describe it as if the data were spatially homogeneous [87]. In principle, a nonlinear version of ICA might solve this problem of distributional heterogeneity [88], but excessive computational requirements for high-dimensional data, and the likelihood of non-unique solutions, cast serious doubt on the practical realization of such nonlinear variants [89–92].

Validation of ICs

Inferring the optimal number of components is difficult to accomplish, since even those eigenimages explaining the smallest variance in the data typically possibly have a statistical structure unlike Gaussian noise. Therefore, the number of ICs to be retrieved is carefully considered (using expertise), combining the need to capture all physiologically interesting components on the one hand and algorithmic and computational constraints on the other hand. The physiologically meaningful task-related ICs are determined post hoc out of the results. However, the lack of a particular order of the resulting ICs means that a given physiological signal source could be expressed in any one of the ICs. Moreover, ICA is not able to determine the absolute energies and sign of the time courses (see 3.3.2.4). This yields a (subjective) interpretation of the resulting ICs or the application of carefully designed criteria to identify the clusters of interest. Several implementations of the latter approach exist:

- Nakada et al. evaluate the correlation coefficient between the time courses of each component and a reference function depicting the task [93]. The components whose associated time courses highly correlate with the paradigm are considered (consistently-) task related, whereas components whose activation is related only partly to the paradigm are classified as transiently task-related components.

- De Martino et al. used a classification based on three descriptive measures: the kurtosis of the map values, the degree of spatial clustering of each map, and the one-lag autocorrelation of each map time series [94, 95]. The kurtosis takes into account the distribution properties of each map intensity, the degree of spatial clustering (after thresholding based on the z-score) was chosen because activation maps usually have a defined spatial structure, and the one-lag autocorrelation was chosen to detect a temporal structure in the maps. The method showed that the simultaneous inspection of these values could reveal potentially meaningful phenomena, because in different tasks the interesting components show similar combinations of these parameters.

- Moritz et al. ranked the independent components of a periodic fMRI complex motor paradigm according to frequency content [96]. Their method identified and ranked the TTR components high, hence separating these from artefacts and confounds.
- McKeown et al. introduced a hybrid procedure [97]. In a first stage, ICA is applied to determine a set of independent components. These ICs are subsequently utilised as a set of regressors for a conventional GLM, which allows to determine a subset of meaningful ICs.

Construction of activation maps

The thresholding operation of each map is usually performed by scaling the intensity values to the z-score. Within each component map, the voxels that contribute significantly to the map are those having a z-score whose absolute value is greater than a threshold. Voxels whose time series are modulated opposite to the time score of the component show a negative z-score. We emphasize once more that the z-score has no statistical significance, but is only used for descriptive purposes. In order to make statistical inferences about these maps, some hypothesis about the distribution of the noise or the signal is needed.

2.4.2.4 Robustness

The algorithmic and statistical reliability of the estimated components were probed by Himberg et al. by running the FastICA algorithm (see section 3.3.3) with different initial conditions or bootstrap samples respectively [98]. Visualizing the cluster structure in the maps showed that the expected components are reliable. On the other hand, components where found whose interpretation was not obvious but showed high reliability, suggesting unknown underlying phenomena.

2.5 Principles and applications of FCM for the analysis of fMRI data sets

2.5.1 Searching for similarities

The analysis of fMRI data involves the identification of regions of activation and the corresponding temporal behaviour associated with a task. It can be assumed that the pattern of activation actually has a structure and can be divided into a few types of activations. Clustering techniques try to classify signals into several groups or clusters according to similarity among these signals, quantified using a distance measure. Each cluster is characterized by its spatial map and the corresponding time course, the cluster centre. Subsequently, cluster centres can be analysed with regard to descriptive parameters such as activation strength and delay. Clustering techniques provide additional information, namely the cluster assignments, i.e. the labels for each of the voxels, according to their similarity. It is therefore possible to isolate zones with similar activation. The clusters derived

during the latter clustering process (unsupervised clustering) can also be used as reference groups for the analysis of similar data sets (supervised clustering).

2.5.1.1 C-means

The standard clustering algorithm for functional imaging is C-means [99, 100]. The clusters are a partition of the data, such that each voxel belongs to exactly one cluster. Fig. 2.4 shows a schematic result of the C-means clustering algorithm for a two-dimensional feature space (i.e. two time samples are taken for each voxel) and three clusters. The clusters are indicated by the circles and the cluster centres by the bold circles.

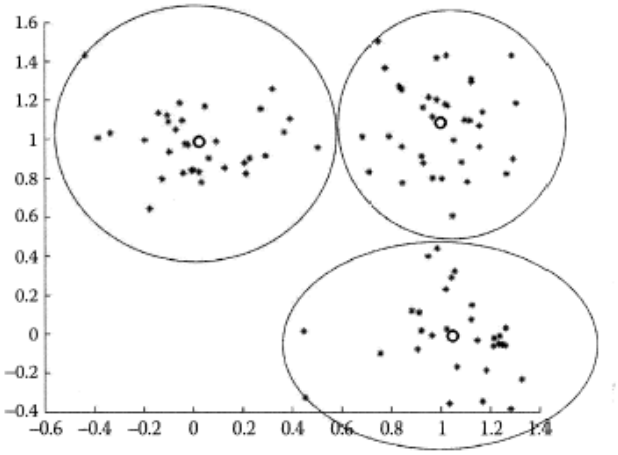

Fig. 2.4: *Example of a C-means clustering result for a two-dimensional data set. Three clusters were searched for. The clusters are indicated by the circles and their centres by the bold circles [101].*

Fig. 2.5 shows typical C-means results, i.e. spatial maps and corresponding time courses, obtained by Goutte et al. in the context of a visual experiment study [102]. The paradigm consisted of a rest period of 20 s of darkness (using a light fixation dot), followed by 10 s of full-field checker board reversing at 8 Hz, and ending by 20 s of darkness. Three runs of 100 images were entered in a C-means algorithm, preceded by a statistical test eliminating most of the non activated voxels. The figure shows the activation maps and corresponding time courses of four of the seven clusters found. The first three (top row and bottom left) are positively correlated with the paradigm and correspond possibly to different regions in the visual cortex, each with different response strength. The fourth cluster is anticorrelated with the stimulus and turned out to reflect the heart

beat. The other three clusters contained voxels that are weakly correlated with the stimulus (not shown).

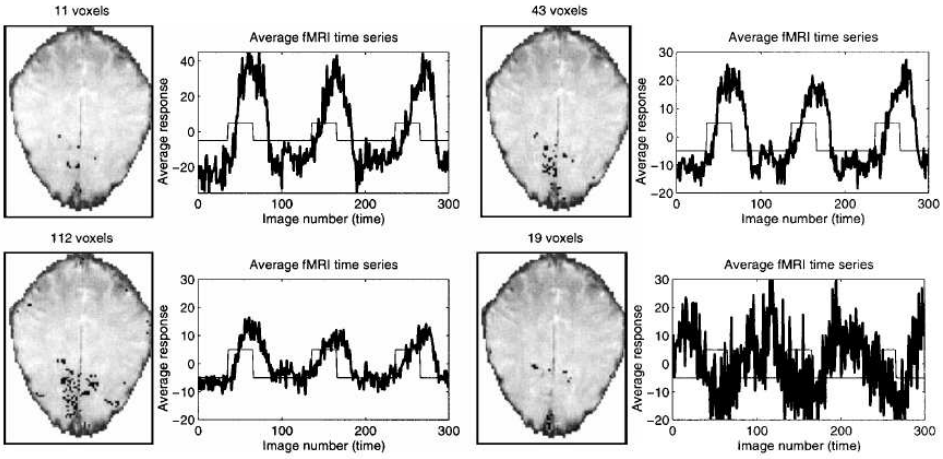

Fig. 2.5: *Example of maps and time courses resulting from a C-means clustering procedure, taken from [102]. The first three clusters (top row and bottom left) are positively correlated with the paradigm and correspond possibly to different regions in the visual cortex, each with different response strength. The fourth cluster is anticorrelated with the stimulus and turned out to reflect the heart beat.*

2.5.1.2 Fuzzy C-means

The C-means algorithm was noticed to lack potentially in reproducibility (see section 2.5.2.4). In order to deal with this shortcoming, the fuzzy clustering method (FCM) was introduced by Zadeh in 1977 [103]. Fuzziness relates to the fact that each voxel belongs to all clusters to a certain degree. The similarity of a voxel to each cluster centre is expressed by a so-called membership to that cluster. FCM takes the latter approach into account during the analysis of the data. It results in a set of clusters (maps and centre time courses) and the memberships of all voxels to each cluster. Moreover, a fuzziness coefficient m is introduced determining the influence of the memberships in the latter procedure (see 2.5.2.2). In 1981, Bezdek presented an iterative algorithm based on the latter approach [104, 105]. This algorithmic implementation proved to be very fast.
Finally, we mention that FCM represents the data both in a nonlinear manner and locally, i.e. the cluster centroids rarely involve the combination of all time courses. The former is opposed to the usual linear approach of PCA and ICA.

The concept of FCM is illustrated in Fig. 2.6, considering a one-dimensional example. Looking at the scatter plot of the observations of variable z (Fig. 2.6a), two

clusters can be identified, called 'A' and 'B'. The C-means approach (Fig. 2.6b) leads to an assignment of each observation to one of the clusters only, whereas the FCM-approach indicates how much every observation belongs to each cluster. The corresponding membership function follows a smooth line.

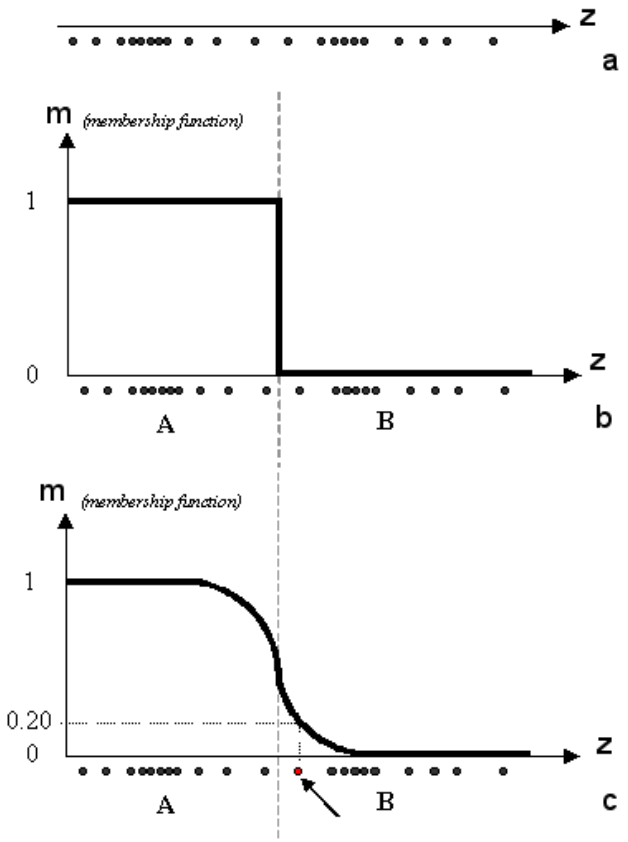

Fig. 2.6: *Difference between FCM and C-means. The C-means approach leads to an assignment of each observation to one of the clusters (A or B) only, whereas the FCM-approach indicates how much every observation belongs to each cluster. As an example, the red coloured observation is considered. Its membership to cluster 'B' is 0.2 and to cluster 'A' 0.8.*

Note that fuzzy algorithms do not give the probability of an individual time course belonging to a particular fuzzy centre, but express the degree to which the time course belongs to the centre. More formally, the probability of a time course being correlated with a particular cluster is an absolute concept, whereas the membership function is always calculated relative to all other clusters.

2.5.2 Strength and shortcomings of FCM

2.5.2.1 Comparison to other techniques

FCM was extensively compared to alternative techniques for the analysis of fMRI data sets in terms of accuracy, robustness, and computation speed.

- Baumgartner et al. compared FCM to correlation analysis and concluded that FCM outperformed for typical low CNR data sets in terms of detection accuracy [106]. In a subsequent study comparing the performance of FCM and PCA, Baumgartner et al. showed that FCM outperformed PCA in the entire CNR range studied in terms of detection accuracy for physiological noise [107]. The results for scanner noise, results were comparable.

- Dimitriadou et al. performed a comparison between several clustering algorithms, such as FCM, C-means, SOM, neural gas, and hierarchical clustering [108]. Their analysis was performed as function of several major design factors, such as noise characteristics, CNR level, and initial number of clusters. In this study, a hybrid data set was generated out of baseline in vivo data and artificially added activation. FCM resulted in a better classification. The authors concluded however that the neural gas algorithm seemed the best choice for non-hierarchical cluster analysis.

- Hierarchical clustering was suggested in case of reasonably small data sets. Lu et al. compared FCM to the region-growing method [44] and to a hybrid method based on split-merge and region-growing techniques [109], conducted on simulated and in vivo data sets. They concluded that, compared to FCM, their hybrid method finds more homogeneous brain regions, i.e. more continuous and fewer scattered clusters were found.

- Meyer-Baese et al. compared (amongst other clustering algorithms) FCM based on deterministic annealing (a hierarchical algorithm monitoring different control parameters: the free energy and entropy) to FastICA and InfoMax [110]. Fuzzy clustering outperformed both ICA algorithms in terms of classification results and sensitivity range of the CTR component, but required a longer processing time.

2.5.2.2 Parameter selection and methodological approaches

The first studies applying FCM to fMRI data sets performed clustering directly on the time-series and applied the Euclidean distance to quantify the similarity among the acquired signals (similar for C-means). Furthermore, they selected a value for method-specific parameters, such as the fuzziness coefficient and the number of clusters, without extensive investigation about their influence on the results. A large amount of studies exist, investigating alternative implementations of distance

measures and extracted features to cluster on. Moreover, they try to determine well-considered settings for the aforementioned method-specific parameters. In this paragraph, we will elaborate on these aspects as to elucidate their importance in this dissertation. The approach concerning the 'number of clusters' however is expounded in section 2.5.2.3 in the context of the cluster validation.

Feature space

Using the raw time series as input is limited by the potentially high dimensional space and the high noise level. Therefore it is better to perform clustering on selected features, in a lower dimensional feature space [111]. Several authors applied clustering to features extracted from the time series, instead of clustering on the raw time series. Goutte et al. investigated the use of two simple features, namely the activation strength and the response delay of the cross-correlation function between the time series and the excitation paradigm, thereby combining hypothesis-and data-driven techniques [112]. In the same study, the authors performed a meta-analysis in a seven-dimensional feature space, using different standard single-voxel analysis techniques, like Student t-test (between rest and activation) and correlation with the paradigm. Simon et al. applied clustering to data from six different task-control pairs (6 Student t-values) collected for each subject in the study [113]. Jahanian et al. presented a feature space based on multiscale decompositions obtained by scalar wavelet and multiwavelet transforms [114].

Alternatively, as is done in this dissertation, clustering can be performed on the averaged (with respect to task onset) time courses. This yields a feature space that still represents time course data of all voxels but with fewer dimensions. Averaging over trials however, may reduce detection sensitivity for TTR activity or activity which is not time-locked to the experimental event (see section 3.5.3).

Distance measure

In order to elaborate on typical distances measures applied when analysing an fMRI data set with FCM, we present some quantities and notations. Let $\boldsymbol{x}_n$ be the time course of voxel n ($n = 1 \dots N$), with N the total number of voxels in a single image. Let $\boldsymbol{v}_c$ be the centre time course of cluster c ($c = 1 \dots C$), with C the number of clusters. Each time course consists of T samples ($t = 1 \dots T$). Therefore, the t-th sample of cluster centre $\boldsymbol{v}_c$ or voxel $\boldsymbol{x}_n$ is represented by $v_{c,t}$ or $x_{n,t}$, respectively. The following measures express the distance between a voxel time course $\boldsymbol{x}_n$ and a cluster centre $\boldsymbol{v}_c$ in a feature space consisting of time samples.

Typically, in FCM, the distance measure is implemented as the Euclidean distance, defined by the L_2-norm:

$$d_{\mathrm{E}}\left(\boldsymbol{x}_n, \boldsymbol{v}_c\right) = \left\|\boldsymbol{x}_n - \boldsymbol{v}_c\right\|^2 = \sqrt{\sum_{t=1}^{T}\left(v_{c,t} - x_{n,t}\right)^2} \quad . \tag{2.1}$$

However, in order to overcome problems related to the data set or the analysis method, other measures were suggested. Bradley et al. proposed to reduce the convergence problem of C-means by using the L_1-norm as a distance measure:

$$d_{\mathrm{L}_1}\left(\boldsymbol{x}_n, \boldsymbol{v}_c\right) = \sum_{t=1}^{T}\left|v_{c,t} - x_{n,t}\right| \quad , \tag{2.2}$$

thereby calling their method C-median [115]. The latter measure is less sensitive to outliers than the traditional Euclidean L_2-norm. Due to the high noise level in fMRI experiments, the results of clustering on raw time series are often unsatisfactory and do not necessarily group data according to the similarity of their pattern of response to the stimulus. This consideration has led Golay et al. [116] and Toft et al. [117] in two independent abstracts to consider a metric based on the correlation between stimulus and time series:

$$d_{\rho}\left(\boldsymbol{x}_n, \boldsymbol{v}_c\right) = \left(\frac{1 - \rho\left(\boldsymbol{x}_n, \boldsymbol{v}_c\right)}{1 + \rho\left(\boldsymbol{x}_n, \boldsymbol{v}_c\right)}\right)^m \quad , \tag{2.3}$$

with ρ Pearson's cross-correlation coefficient and m the membership.

The Mahalanobis distance [118] is defined as:

$$d_{\mathrm{M}}\left(\boldsymbol{x}_n, \boldsymbol{v}_c\right) = \left(\boldsymbol{x}_n - \boldsymbol{v}_c\right)\ \boldsymbol{C}_c{}^{-1}\ \left(\boldsymbol{x}_n - \boldsymbol{v}_c\right)^T \quad . \tag{2.4}$$

Instead of treating all voxels $\boldsymbol{x}_n$ equally when calculating the distance to the cluster centre $\boldsymbol{v}_c$, the Mahalanobis distance weights the differences by the range of variability, described by cluster's covariance matrix $\boldsymbol{C}_c$, in the direction of the voxel [102]. The Euclidean distance does not take into account the shape of the cluster, i.e. it assumes a spherical shape, corresponding to a covariance matrix $\boldsymbol{C}_c$ with ones on the main diagonal and zeros elsewhere. The matrix inversion however results in longer computation times and may lead to singularity problems when estimating the covariance matrix during the iterative process.

Fuzziness coefficient

The fuzziness coefficient m theoretically lies between 1 and plus infinity. When m equals 1, hard partitioning is performed and FCM is similar to C-means. When m tends to infinity, the membership values tend to $1/C$, i.e. each voxel is assigned to all clusters to the same degree. The ideal value of the fuzziness coefficient is problem-dependent and has to be determined a priori. However, Fadili et al. noticed that for values of m larger than 3, it was difficult to distinguish between the non-activated and activated voxels [119]. Common values for fMRI data set

analysis range from 1.3, which was experimentally determined [120] to two, which was determined using ROC methodology [119].

2.5.2.3 Interpretation of the results

The interpretation of the results concerns the determination of maps of significant activation and their corresponding time courses out of the FCM results. Preliminary, however, the validation of the results, i.e. the selection of physiologically interesting clusters has to be accomplished. Both aspects are elucidated in this chapter. However, the interpretation is severely hampered by the large amount of inactivated voxels, a phenomenon known as the 'ill-balanced data problem'. The latter is therefore expounded in detail as well.

Ill-balanced data problem

FCM is very sensitive to the large amount of inactivated voxels [119]. The clusters found can be biased by voxels containing movement-related artefacts or large-vein contributions. This 'ill-balanced data problem' is usually alleviated by an adequate preprocessing technique reducing the number of voxels to be analysed. Several approaches exist to determine such a subset:

- A first approach is to distinguish between noisy voxels and potentially interesting ones by analysing the temporal structure in the measured data. Fadili et al. quantified the latter using autocorrelation [119]. Baumgartner et al. used one-lag autocorrelation, Spearman's rank-order based correlation, and measures of negentropy and kurtosis [121]. Jarmasz and Somorjai defined a statistic that measures the departure from flatness of the power spectrum of the time courses [122].

- Furthermore, 'classical' analysis techniques can be applied to decompose an fMRI data set in two classes only as to separate signal and noise. Yao et al. for example performed an SVD decomposition of the delay-correlation matrix as to separate signal subspace from noise subspace [123].

- Finally, functional information can be used to determine a region of interest (ROI). Alternatively anatomical information can be used to determine a subset, since functional activity is restricted to those voxels that lie within the cortex (gray matter) [119, 124, 125].

Cluster validity problem

FCM requires an a priori definition of the number of clusters. However, the exact number is mostly not known. Therefore it is carefully considered (using expertise), combining the needs to capture all physiologically interesting components and to avoid the further subdivision of functionally uniform clusters. The physiologically meaningful clusters must therefore be determined post hoc out of the results. However, the lack of a particular order of the resulting clusters means that a given

physiological signal source could be expressed in any one of the clusters. This yields a (subjective) interpretation of the resulting clusters or the application of carefully designed criteria to identify the clusters of interest. In FCM literature, this issue is known as the 'cluster validity problem'. Several approaches were proposed to deal with this problem.

- Some authors introduced heuristics as to estimate the optimal number of clusters. The result however is often problem-dependent. Fadili et al. introduced a criterion based on compactness, separation, fuzzy intersection and fuzzy union to deduce a quantitative measure to compare partitions with different number of clusters [126]. Jahanian et al. introduced an ROC (see section 2.6) based technique to determine the number of clusters [127].
- Another approach consists in the validation of the resulting clusters. Möller et al. proposed an approach consisting of two steps [128]. Firstly, a sequence of partitions with increasing number of clusters is created from the data. Secondly, a validity index is applied to select the partition exhibiting the clearest indication of an existing structure in the data. Baumgartner et al. used a resampling technique to validate the results of an FCM analysis and computed the statistical significance for each voxel to belong to a cluster [129]. Auffermann et al. assessed the statistical significance associated with partitioning one cluster into two clusters or the inverse problem of combining two clusters into one. Their method is based on Fisher's linear discriminant and the bootstrap and was demonstrated for the SOM clustering algorithm [130].

- Alternatively, clustering algorithms have been presented that do not need the number of clusters as prior information, such as hierarchical clustering [102, 131]. This technique consists of two parts: the division and merging part. During the first part, clusters are iteratively divided into smaller entities, using the C-means clustering algorithm. Therefore, two clusters are considered in each iteration step. The procedure is halted when there is no significant structure left in the data. During the merging part, similar clusters (defined in terms of a distance measure, such as the Euclidean distance) are iteratively re-combined. The procedure is stopped when there is a sudden increase of the distances for the remaining cluster pairs. Another approach consists in dynamical clustering. This technique avoids to fix the number of clusters a priori by generating and annihilating cluster centres dynamically, i.e. during the data fitting process. Algorithms implementing the latter approach are ISODATA [132] and Dynamical Clustering Analysis (DCA) [133]. Baune et al. showed that compared to C-means, DCA led to much better reproducibility of results, at the expense of computation time however.
- Finally, a hybrid approach is often proposed. In a first step, hierarchical clustering may be applied as to determine an estimate of the number of

clusters and their centres. These results are subsequently used to initialise the actual FCM procedure. An example of this approach (for segmentation purposes) is described in [134].

It is worth emphasizing that the lack of a particular order of the resulting clusters and the corresponding necessity of a validation are similar for the independent components resulting from an ICA analysis. FCM however, is able to determine the absolute energies and the sign of the cluster centroids, as opposed to ICA.

Construction of activation maps

The C-means algorithm results in a hard partitioning (crisp clustering), i.e. each voxel belongs to one cluster only. The physiologically meaningful clusters are therefore converted to activation maps without further manipulation.
FCM results consist of membership maps of all voxels to the clusters and their corresponding centroids, representing the average time course of the cluster. Within each cluster map, the voxels that contribute significantly to the map are selected by imposing a (user-defined) threshold on the membership. Several authors suggested 0.8 as a threshold value ($m > 0.8$) as to impose a high degree of similarity of the remaining voxels to the cluster centroids [135, 136]. In our research, a threshold of 0.5 is applied, as to assure that each voxel is assigned to at most one cluster.

2.5.2.4 Accuracy, robustness, and convergence speed

Scarth et al. were the first to apply FCM on fMRI data [120]. In a study with simulated and real fMRI data of motor and cognitive tasks, they demonstrated that the technique was able to identify both the regions and the temporal nature of the functional activations. The influence of higher field strengths on accuracy was investigated in several studies. In an experiment applying a field strength of 3 T, Barth et al. [137] were able to separate large vessel from small vessel activation, whereas Windischberger et al. [138] identified and locally separated various artefacts.

The robustness of the applied clustering techniques was a key feature in the shift from C-means to FCM. The robustness of C-means is poor, because of strong dependence on the initialization of the cluster centres especially for high-dimensional fMRI data sets (i.e. with hundreds of sampling points in time). This is due to the fact that the gradient-descent performed by the C-means algorithm is a local minimizer. The latter frequently fails to find the global minimum for high-dimensional data sets. Fuzzy clustering on the other hand proved to be a robust method for the extraction of localized haemodynamic responses from the microvasculature. [135, 139].

Convergence of C-means and FCM is usually very fast. This is an important

feature in view of the large and complex data sets to be analysed, as convincingly argued in many studies [140, 141].

2.6 ROC analysis

In this section, the Receiver Operating Characteristics (ROC) technique to assess the inherent accuracy of a detection procedure is elucidated. The assessment of the detection capability inherent to a analysis method implies measuring separately the detection accuracy for actually positive (i.e. activated) and actually negative (i.e. not activated) voxels. This can only be accomplished if the ground truth is available. For that purpose, a synthetic data set is constructed, in which mimicked activation is added to a real background data set (i.e. real noise). The assessment is then expressed in terms of a pair of indices: the 'sensitivity and the 'specificity. The sensitivity is the fraction of voxels actually activated that is correctly diagnosed as activated. This fraction is also called the 'true-positive rate or TPR. The specificity on the other hand is the fraction of voxels actually not activated that is correctly diagnosed as not activated. As an alternative to the latter, one usually expresses the fraction of voxels actually not activated that is wrongly diagnosed as activated. This fraction is called the 'false-positive rate or FPR.

These two figures of merit, TPR and FPR are not fixed values of the detection procedure however, but are dependent on a 'rating parameter', expressing the applied sensitivity of the detection procedure [142–144]. However, for the assessment of the detection procedure, the extraction of a single figure of merit to assess the procedure is desirable. This is accomplished in two steps.

Firstly, an ROC curve is established. As shown in Fig. 2.7, an ROC is a continuous plot of TPR versus FPR in a unit square for a range of threshold values of the rating parameter. In this dissertation, the ROC methodology is applied on the FCM method. FCM results in a set of memberships of each voxel to all clusters, expressing the degree to which a voxel is similar to the cluster centre. Therefore, such a membership acts as a rating parameter of FCM. The decision criterion thus corresponds to a specific threshold on the membership [[119, 139]. A low threshold value (close to 0) assesses many imaging voxels as being activated, resulting in large values for TPR and FPR. A high threshold value (close to 1) assesses few imaging voxels as being activated, resulting in low values for TPR and FPR).
It is worth mentioning that the basic assumption for ROC curve fitting (i.e. turning discretely calculated data into a continuous plot) is that the underlying data for truly positive (i.e. activation) and truly negative (i.e. no activation) trials form a binormal distribution. However, binormal ROC methods are also useful when the underlying distributions of test results for truly positive and truly negative trials are unknown and/or non-Gaussian. It is only necessary that the rating parameter can be transformed to create an approximately binormal distribution [145]. These conditions are satisfied for the FCM method analysing fMRI data,

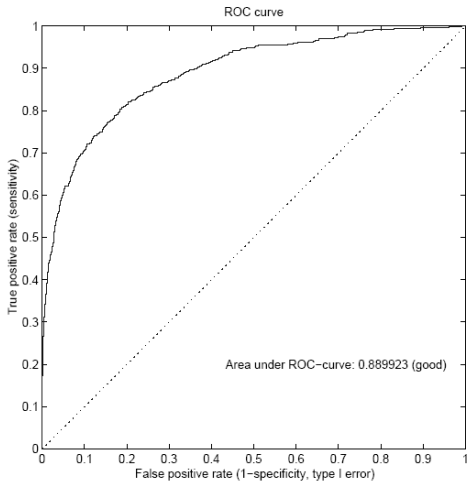

Fig. 2.7: *Typical example of an ROC curve. The solid line shows the FPR versus the TPR as function of a rating parameter of the detection procedure. The dotted line represents a random classification. The area under the solid line (0.89) is a measure for the assessment of the detection procedure.*

making ROC methodology applicable.

The second step consists in extracting the best single parameter out of the ROC curve for characterizing the methods detection accuracy. Because ROC curves indicating better decision performance lie higher in the unit square, the area under the fitted ROC curve, commonly denoted as A_z, is widely for this purpose. In fMRI experiments however, the ratio of false activations is much smaller than the ratio of real activations. Indeed, the number of activated voxels involved in a particular stimulation is much smaller than the number of inactivated voxels (ill-balanced data problem). Therefore, A_z is not desirable as a figure of merit for ROC analysis of fMRI data sets. As an alternative Skudlarski et al. [146] introduced the mean of the ROC curve over the limited range of false-positive ratio between 0 and (the somewhat arbitrary value of) 0.1 as a measure. By limiting the value of merit to low, but realistic false-positive rates (high thresholds only) we limit the scope of our analysis to the cases that are of primary interest in fMRI.

2.7 Conclusions

In the analysis of BOLD fMRI data, the estimation of the spatial layout of task-related brain activation was originally based on hypothesis-driven methods. This category of methods assumes a spatially-invariant model of the HR function. It was illustrated that this assumption is not optimal for the analysis of complex tasks.

Indeed, these tasks normally involve the activation of extended networks of brain regions with widely different HRs. Additionally, within each region, a substantial degree of experimentally-induced trial-by-trial variability is to be expected. Moreover, hypothesis-driven methods are mostly univariate, thereby ignoring interactions between voxels.

A complementary approach to estimate the spatio-temporal pattern of brain activation was introduced: data-driven methods, such as spatial Independent Component Analysis and the Fuzzy Clustering Method. In both these methods, a representation of the data is obtained by decomposing the data into a set of spatio-temporal modes, without strong a priori assumptions about the temporal profile of the effects of interest. FCM is applied to decompose the data set into a set of *time* patterns with associated spatial maps (i.e. a temporal view on the data is adopted). In case of ICA however, a spatial view on the data is adopted (sICA). The data set is decomposed into a set of *spatial* patterns of activation with associated time courses.

Both techniques are successful in retrieving complex patterns of brain activation often left undetected by other methods or hard to predict beforehand. However, the correct validation of the results and a well-considered selection of method-specific parameters and approaches are subjected to the complexity of the data set and method-specific features. In this dissertation, dealing with the dissection of the distinct stages of a cognitive task, these shortcomings may severely impede a successful application of the methods. Therefore, in chapter 3, we will elaborate on the characteristics of specific algorithms implementing FCM and sICA and thoroughly investigate the influence of their parameters. The guideline is to introduce approaches based on both techniques as to improve the detection accuracy, consistency and robustness when analysing complex data sets. In order to provide data sets allowing to evaluate these criteria, a newly devised mental imagery experiment will be introduced.

Bibliography

Bibliography

[1] S Ogawa et al. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci USA*, 89(13):5951–5955, Jul 1992.

[2] K K Kwong et al. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc Natl Acad Sci USA*, 89 (12):5675–5679, Jun 1992.

[3] P A Bandettini et al. Time course EPI of human brain function during task activation. *Magn Reson Med*, 25(2):390–398, Jun 1992.

[4] K J Friston et al. Characterizing dynamic brain responses with fMRI: a multivariate approach. *Neuroimage*, 2(2):166–172, Jun 1995b.

[5] S Gold et al. Functional MRI statistical software packages: a comparative analysis. *Hum Brain Mapp*, 6(2):73–84, 1998.

[6] K K Kwong. Functional magnetic resonance imaging with echo planar imaging. *Magn Reson Q*, 11(1):1–20, Mar 1995.

[7] G R Norman and D L Steiner. *Biostatistics The bare essentials*. B C Decker Inc., Hamilton London, 2000.

[8] P A Bandettini et al. Processing strategies for time-course data sets in functional MRI of the human brain. *Magn Reson Med*, 30(2):161–173, Aug 1993.

[9] G K Aguirre et al. The variability of human, BOLD hemodynamic responses. *Neuroimage*, 8(4):360–369, Nov 1998.

[11] G M Boynton et al. Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci*, 16(13):4207–4221, Jul 1996.

[10] A M Dale and R L Buckner. Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, 5:329–340, 1997.

[12] A L Vasquez and D C Noll. Nonlinear aspects of the BOLD response in functional MRI. *Neuroimage*, 7:108–118, 1998.

[13] G H Glover. Deconvolution of impulse response in event-related fMRI. *Neuroimage*, 9:416–429, 1999.

[14] M D Robson et al. Measurements of the temporal fMRI response of the human auditory cortex to trains of tones. *Neuroimage*, 7(3):185–198, Apr 1998.

[15] K J Friston et al. Analysis of functional MRI time-series. *Human Brain Mapping*, 1:153–171, 1994.

[16] G K Aguirre et al. Empirical analyses of BOLD fMRI statistics. II. Spatially smoothed data collected under null-hypothesis and experimental conditions. *Neuroimage*, 5(3):199–212, Apr 1997.

[17] N Lange and S L Zeger. Non-linear Fourier time series analysis for human brain mapping by functional Magnetic Resonance Imaging . *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(1):1–29, 1997.

[18] J C Rajapakse et al. Modeling hemodynamic response for analysis of functional MRI time-series. *Hum Brain Mapp*, 6(4):283–300, 1998.

[19] C H Liao et al. Estimating the delay of the fMRI response. *Neuroimage*, 16 (3 Pt 1):593–606, Jul 2002.

[20] M A Burock and A M Dale. Estimation and detection of event-related fMRI signals with temporally correlated noise: a statistically efficient and unbiased approach. *Hum Brain Mapp*, 11(4):249–260, Dec 2000.

[21] C R Genovese. A Bayesian time-course model for functional magnetic resonance imaging data. *Journal of the American Statistical Association*, 95: 691–719, 2000.

[22] K J Friston et al. Statistical parametric maps in functional imaging: A general linear approach . *Human Brain Mapping*, 2(4):189–210, 1995a.

[23] K J Friston et al. Comparing functional (PET) images: the assessment of significant change. *J Cereb Blood Flow Metab*, 11(4):690–699, Jul 1991.

[24] V P Clark et al. fMRI study of face perception and memory using random stimulus sequences. *J Neurophysiol*, 79(6):3257–3265, Jun 1998.

[25] S D Forman et al. Improving assessment of significant activation in fMRI: Useof a cluster-size threshold. *Magn Reson Med*, 33:636–647, 1995.

[27] C R Genovese et al. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878, Apr 2002.

[26] Y Hochberg and Y Benjamini. More powerful procedures for multiple significance testing. *Stat Med*, 9(7):811–818, Jul 1990.

[28] D B Rowe and R G Hoffmann. Multivariate statistical analysis in fMRI. *IEEE Eng Med Biol Mag*, 25(2):60–64, Mar 2006.

[29] L K Hansen. Multivariate strategies in functional magnetic resonance imaging. *Brain Lang*, Jan 2007.

[30] C G Phillips et al. Localization of function in the cerebral cortex. Past, present and future. *Brain*, 107:327–361, Mar 1984.

[31] D J Quint and J L Gilmore. Alexia without agraphia. *Neuroradiology*, 34 (3):210–214, 1992.

[32] O Friman et al. Detection of neural activity in functional MRI using canonical correlation analysis. *Magn Reson Med*, 45(2):323–330, Feb 2001.

[33] R Nandy and D Cordes. Improving the spatial specificity of canonical correlation analysis in fMRI. *Magn Reson Med*, 52(4):947–952, Oct 2004.

[34] PC Fletcher et al. Is multivariate analysis of PET data more revealing than the univariate approach? Evidence from a study of episodic memory retrieval.

[35] S Mallat. A theory of multiresolution signal decomposition: The Wavelet representation. *IEEE Trans Pattern Anal Mach Intell*, 11:674–693, 1989.

[36] M J Brammer. Multidimensional wavelet analysis of functional magnetic resonance images. *Hum Brain Mapp*, 6(5-6):378–382, 1998.

[37] E Bullmore et al. Wavelets and functional magnetic resonance imaging of the human brain. *Neuroimage*, 23 Suppl 1:234–249, 2004.

[38] O Friman et al. Exploratory fMRI analysis by autocorrelation maximization. *Neuroimage*, 16(2):454–464, Jun 2002.

[39] J A Hartigan. *Clustering algorithms*. John Wiley and Sons, 1975.

[40] T Hofmann and J Buhmann. Pairwise data clustering by deterministic annealing. 1996.

[41] T M Martinetz and K J Schulten. A neural gas network learns topologies. In T Kohonen et al., editors, *Artificial Neural Networks*, pages 397–402, North Holland, Amsterdam, 1991.

[42] A Wismuller et al. Model-free functional MRI analysis based on unsupervised clustering. *J Biomed Inform*, 37(1):10–18, Feb 2004.

[43] Jia-Hong Gao and Seong-Hwan Yee. Iterative temporal clustering analysis for the detection of multiple response peaks in fMRI. *Magn Reson Imaging*, 21(1):51–53, Jan 2003.

[44] Y Lu et al. A split-merge-region based region-growing method for fMRI activation detection. *Human Brain Mapping*, 22(4):271–279, 2004.

[45] T Kohonen. Automatic formation of topological maps of patterns in a self-organizing system. In E Oja and O Simula, editors, *Proceedings of 2SCIA, Scand. Conference on Image Analysis*, pages 214–220, Helsinki, Finland, 1981. Suomen Hahmontunnistustutkimuksen Seura r.y.

[46] H Fisher and J Hennig. Clustering of functional mr data. In *Proceedings ISMRM 4rd Annual Meeting*, volume 96, pages 1179–1183, 1996.

[47] S Lai and G H Glover. Detection of BOLD fMRI signals using complex data . page 1671, 1999.

[48] F Y Nan and Nowak R D. Generalized likelihood ratio detection for fMRI using complex data. *IEEE Trans. Med. Imag.*, 18(4):320–329, 1999.

[49] D B Rowe. Modeling both the magnitude and phase of complex-valued fMRI data. *Neuroimage*, 25(4):1310–1324, May 2005.

[50] K Pearson. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, 2:559–572, 1901.

[51] H Hotelling. Analysis of a complex of statistical variables into principle components. *Journal of educational psychology*, 24:417–441 and 498–520, 1933.

[52] K J Friston et al. Functional connectivity: the principal-component analysis of large (PET) data sets. *J Cereb Blood Flow Metab*, 13(1):5–14, Jan 1993.

[53] J J Sychra et al. Synthetic images by subspace transforms. I. Principal components images and related filters. *Med Phys*, 21(2):193–201, Feb 1994.

[54] ET Bullmore et al. Functional magnetic resonance image analysis of a large-scale neurocognitive network. *Neuroimage*, 4:16–33, 1996.

[55] P P Mitra et al. The nature of spatiotemporal changes in cerebral hemodynamics as manifested in fMRI . *Magn Reson Med*, 37(4):511–518, 1997.

[56] S Dodel et al. Localization of brain activity-blind separation for fMRI data. *Neurocomputing*, 32-33:701–708, 2000.

[57] A Hyvärinen and E Oja. Independent component analysis: algorithms and applications. *Neural Netw*, 13(4-5):411–430, May 2000.

[58] J Hérault and B Ans. Circuits neuronaux *à* synapses modifiables: décodage de messages composites par apprentissage non supervisé. *C.-R. de l'Académie des Sciences*, 299(III-13):525–528, 1984.

[59] B Ans et al. Adaptive neural architectures: detection of primitives. In *Proceedings of COGNITIVA '85*, pages 593–597, Paris, France, 1985.

[60] P Comon. Separation of stochastic processes. In *Proc Workshop on higher-order spectral analysis*, pages 174–179, Vale, Colorado, 1989.

[61] J F Cardoso. Blind identification of independent signals. In *Proc Workshop on higher-order spectral analysis*, Vale, Colorado, 1989.

[62] C Jutten and J Hérault. Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture. *Signal Proc*, 24:1–10, 1991.

[63] M Gaeta and JL Lacoume. Source separation without prior knowledge: the maximum likelihood solution. In *Proceedings EUSPICO*, pages 621–624, 1990.

[64] DT Pham et al. Separation of a mixture of independent sources through a maximum likelihood approach. In *Proceedings EUSPICO*, pages 771–774, 1992.

[65] P Comon. Independent component analysis - a new concept? *Signal Proc*, 36:287–314, 1994.

[66] A J Bell and T J Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput*, 7(6):1129–1159, Nov 1995.

[67] S Makeig et al. Blind separation of auditory event-related brain responses into independent components. *Proc Natl Acad Sci USA*, 94(20):10979–10984, Sep 1997.

[68] M J McKeown et al. Spatially independent activity patterns in functional MRI data during the stroop color-naming task. *Proc Natl Acad Sci USA*, 95 (3):803–810, Feb 1998a.

[69] A Hyvärinen and E Oja. A fast fixed-point algorithm for independent component analysis. *Neural Comput*, 9(7):1483–1492, 1997.

[70] F Esposito et al. Spatial independent component analysis of functional MRI time-series: to what extent do results depend on the algorithm used? *Human Brain Mapping*, 1:146–157, 2002.

[71] V D Calhoun et al. Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Hum Brain Mapp*, 13(1):43–53, May 2001.

[72] M J McKeown et al. Analysis of fMRI data by blind separation into independent spatial components. *Hum Brain Mapp*, 6(3):160–188, 1998b.

[73] B B Biswal and J L Ulmer. Blind source separation of multiple signal sources of fMRI data sets using independent component analysis. *J Comput Assist Tomogr*, 23(2):265–271, 1999.

[74] K S Peterson et al. On the independent components of functional neuroimages. In P Pajunen and J Karhunen, editors, *Proceedings ICA2000*, Helsinki. Espoo Finland, 2000. Otamedia.

[75] N Lange et al. Plurality and resemblance in fMRI data analysis. *Neuroimage*, 10(3 Pt 1):282–303, Sep 1999.

[76] S Zeki et al. The processing of kinetic contours in the brain. *Cereb Cortex*, 13:189–202, 2003.

[77] M Castelo-Branco et al. Activity patterns in human motion sensitive areas depend on the interpretation of global motion. In *Proc Natl Acad Sci USA*, volume 99, pages 13914–13919, 2002.

[78] A Bartels and S Zeki. The architecture of the colour centre in the human visual brain: new results and a review. *Eur J Neurosci*, 12:172–193, 2000.

[79] V D Calhoun et al. Different activation dynamics in multiple neural systems during simulated driving. *Human Brain Mapping*, 16:158–167, 2002.

[80] E Seifritz et al. Spatiotemporal pattern of neural processing in the human auditory cortex. *Science*, 297:1706–1708, 2002.

[81] H Gu et al. Mapping transient, randomly occuring neurophysiological events using independent component analysis. *Neuroimage*, 14:1432–1443, 2001.

[82] P R Montague et al. Hyperscanning: simultaneous fMRI during linked social interactions. *Neuroimage*, 16:1159–1164, 2002.

[83] C G Thomas et al. Noise reduction in bold-based fMRI using component analysis. *Neuroimage*, 17:1521–1537, 2002.

[84] V Perlbarg et al. CORSICA: correction of structured noise in fMRI by automatic identification of ICA components. *Magnetic Resonance Imaging*, 25:35–46, 2007.

[85] M J McKeown et al. Independent component analysis of functional MRI: what is signal and what is noise? *Curr Opin Neurobiol*, 13(5):620–629, Oct 2003.

[86] M J McKeown and T J Sejnowski. Independent component analysis of fMRI data: examining the assumptions. *Hum Brain Mapp*, 6(5-6):368–372, 1998.

[87] J Karhunen et al. Local linear independent component analysis based on clustering. *Int J Neural Syst*, 10(6):439–451, Dec 2000.

[88] L Parra et al. Statistical independence and novelty detection with information-preserving nonlinear maps. *Neural Computation*, 8:260–269, 1996.

[89] F H Lin et al. Comparison of orthogonal and independent component analysis on dimension-reduced fMRI data in partial least squares framework. In *Proc. Intl. Soc. Mag. Reson. Med.*, volume 9, pages 1730–..., Glasgow, 2001.

[90] H Yang et al. Information theoretic approach to blind separation of sources in non-linear mixtures. *Signal Processing*, 64:291–300, 1998.

[91] A Taleb and C Jutten. Nonlinear source separation: the post-nonlinear mixtures. In *Proc. 1997 European Symp. on artifical neural networks*, pages 279–284, Bruges, Belgium, 1997.

[92] K J Friston et al. Nonlinear PCA: characterizing interactions between modes of brain activity. *Phil Trans R Soc Lond Biological Sciences*, 355, 2000.

[93] T Nakada et al. Independent component-cross correlation-sequential epoch (ICS) analysis od high field fMRI time series: direct visualisation of the dual representation of the primary motor cortex in human . *Neurosci Res*, 37: 237–244, 2000.

[94] E Formisano et al. Spatial independent component analysis of functional magnetic resonance imaging time-series: characterization of the cortical components. *Neurocomputing*, 49:241–254, 2002.

[95] F De Martino et al. Classification of fMRI independent components using IC-fingerprints and support vector machine classifiers. *Neuroimage*, 34(1): 177–194, Jan 2007.

[96] C H Moritz et al. Power spectrum ranked independent component analysis of a periodic fMRI complex motor paradigm . *Human Brain Mapping*, 18: 111–122, 2003.

[97] M J McKeown. Detection of consistently task-related activations in fMRI data with hybrid independent component analysis. *Neuroimage*, 11(1):24–35, Jan 2000.

[98] J Himberg et al. Validating the independent components of neuroimaging time series via clustering and visualization. *Neuroimage*, 22(3):1214–1222, 2004.

[99] J MacQueen. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1:281–297, 1967.

[100] J A Hartigan and M A Wong. Algorithm AS136. A K-means algorithm. *Appl. Stat.*, 28:100–108, 1979.

[101] M F Santarelli et al. *Advanced Image Processing in Magnetic Resonance Imaging*. Taylor and Francis, 2005.

[102] C Goutte et al. On clustering fMRI time series. *Neuroimage*, 9(3):298–310, Mar 1999.

[103] L A Zadeh. *Fuzzy Sets and their Application to Pattern Classification and Clustering Analysis*. Academic Press, New-York, 1977.

[104] J C Bezdek. *Pattern recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New-York, 1981.

[105] J C Bezdek. FCM: the fuzzy C-means algorithm. *Computers and Geosciences*, 10:191–203, 1984.

[106] R Baumgartner et al. Quantification in functional magnetic resonance imaging: fuzzy clustering vs. correlation analysis. *Magn Reson Imaging*, 16(2): 115–125, 1998.

[107] R Baumgartner et al. Comparison of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis. *Magn Reson Imaging*, 18(1):89–94, Jan 2000a.

[108] E Dimitriadou et al. A quantitative comparison of functional MRI cluster analysis. *Artif Intell Med*, 31(1):57–71, May 2004.

[109] Y Lu et al. A split-merge-region based region-growing method for fMRI activation detection. *Human Brain Mapping*, 22(4):271–279, 2004.

[110] A Meyer-Baese et al. Comparison of two exploratory data analysis methods for fMRI: unsupervised clustering versus independent component analysis. *IEEE Trans Inf Technol Biomed*, 8(3):387–398, Sep 2004.

[111] C Goutte et al. Detection of space-time analysis of fMRI by feature space clustering. In A. Evans T Paus, A Gjedde, editor, *Fourth International Conference on Functional Mapping of the Human Brain*, volume 3. Academic Press, 1998.

[112] C Goutte et al. Feature-space clustering for fMRI meta-analysis. *Hum Brain Mapp*, 13(3):165–183, Jul 2001.

[113] O Simon et al. Automatized clustering and functional geometry of human parietofrontal networks for language, space, and number. *Neuroimage*, 23 (3):1192–1202, Nov 2004.

[114] H Jahanian et al. Functional magnetic resonance imaging activation detection: fuzzy cluster analysis in wavelet and multiwavelet domains. *J Magn Reson Imaging*, 22(3):381–389, Sep 2005.

[115] P S Bradley et al. *Clustering via concave minimization*, pages 368–374. MIT Press, Cambridge, MA, 1997.

[116] X Golay et al. Fuzzy membership versus probability in cross correlation based fuzzy clustering of fMRI data. In Friberg et al., editor, *Third International Conference on Functional Mapping of the Human Brain*, volume 3, page S481, 1997.

[117] P Toft et al. On clustering fMRI time series. In Friberg et al., editor, *Third International Conference on Functional Mapping of the Human Brain*, volume 3, page S456, 1997.

[118] P C Mahalanobis. On generalized distance in statistics. In *Proceedings National Institute of Sciences India*, volume 12, pages 49–55, 1936.

[119] M J Fadili et al. A multistep unsupervised fuzzy clustering analysis of fMRI time series. *Human Brain Mapping*, 10:160–178, 2000.

[120] G Scarth et al. Detection of novelty in functional images using fuzzy clustering. In *Proceedings of the 3rd Meeting ISMRM*, pages 238–..., Nice, France, 1995.

[121] R Baumgartner et al. Novelty indices: indentifiers of potentially interesting time-courses in functional MRI data. *Magn Reson Imaging*, 18:845–850, 2000c.

[122] M Jarmasz and R L Somorjai. Exploring regions of interest with cluster analysis (EROICA) using a spectral peak statistic for selecting and testing the significance of fMRI activation time-series. *Artif Intell Med*, 25(1):45–67, May 2002.

[123] D Yao and N Rao. A delay subspace decomposition algorithm of EEG source location. *Acta Electron. Sinica*, 28(4):135–138, 2000.

[124] R Goebel and W Singer. Cortical surface-based statistical analysis of functional magnetic resonance imaging data. *Neuroimage*, 9:64, 1999.

[125] E Formisano et al. Cortex-based independent component analysis of fMRI time series. *Magn Reson Imaging*, 22(10):1493–1504, Dec 2004.

[126] M J Fadili et al. On the number of clusters and the fuzziness index for unsupervised FCA application to BOLD fMRI time series. *Med Image Anal*, 5(1):55–67, 2001.

[127] H Jahanian et al. Roc-based determination of number of clusters for fMRI activation detection. In *Proceedings of SPIE*, volume 5370, 2004.

[128] U Moller et al. How to avoid spurious cluster validation? A methodological investigation on simulated and fMRI data. *Neuroimage*, 17(1):431–446, Sep 2002.

[129] R Baumgartner et al. Resampling as a cluster validation technique in fMRI. *J Magn Reson Imaging*, 11(2):228–231, Feb 2000b.

[130] W F Auffermann et al. Cluster significance testing using the bootstrap. *Neuroimage*, 17(2):583–591, Oct 2002.

[131] P Filzmoser et al. A hierarchical clustering method for analyzing functional MR images. *Magn Reson Imaging*, 17(6):817–826, Jul 1999.

[132] G H Ball and D J Hall. A clustering technique for summarizing multivariate data. *Behav Sci*, 12(2):153–155, Mar 1967.

[133] A Baune et al. Dynamical cluster analysis of cortical fMRI activation. *Neuroimage*, 9(5):477–489, May 1999.

[134] M J Kwon et al. Hierarchical Fuzzy segmentation of Brain MR images . *International Journal of imaging systems and technology*, 13:115–125, 2003.

[135] R Baumgartner et al. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part I: reproducibility. *J Magn Reson Imaging*, 7 (6):1094–1101, Nov 1997.

[136] X Golay et al. A new correlation-based fuzzy logic clustering algorithm for fMRI. *Magn Reson Med*, 40(2):249–260, Aug 1998.

[137] M Barth et al. Characterization of BOLD activation in multi-echo fMRI data using fuzzy cluster analysis and a comparison with quantitative modeling. *NMR Biomed*, 14(7-8):484–489, Nov 2001.

[138] C Windischberger et al. Fuzzy cluster analysis of high-field functional MRI data. *Artif Intell Med*, 29(3):203–223, Nov 2003.

[139] E Moser et al. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part II: quantification. *J Magn Reson Imaging*, 7(6): 1102–1108, Nov 1997.

[140] P J Huber. Massive datasets workshop: four years after. *J Computational and graphical statistics*, 8, 1999.

[141] E J Wegman. Huge data sets and the frontiers of computational feasibility. *J Computational and graphical statistics*, 4, 1995.

[142] R T Constable et al. An ROC approach for evaluating functional brain MR imaging and postprocessing protocols . *Magnetic Resonance in Medicine*, 34 (1):57–64, 1995.

[143] J A Sorensen and X Wang. Methods for Evaluation of fMRI Techniques. *Magnetic Resonance in Medicine*, 36:737–744, 1996.

[144] J A Swets. ROC analysis applied to the evaluation of medical imaging techniques. *Investigative Radiology*, 14(2):109–121, 1979.

[145] C E Metz. ROC methodology in radiologic imaging. *Investigative Radiology*, 21(9):720–733, 1986.

[146] P Skudlarski et al. ROC analysis of statistical methods used in functional MRI: individual subjects. *Neuroimage*, 9(3):311–329, 1999.

Part II

FCM AND SPATIAL ICA BASED APPROACHES AND TECHNIQUES FOR THE DISSECTION OF THE PROCESSING STAGES OF A COGNITIVE TASK

Chapter 3

FCM and sICA based approaches and techniques for the analysis of time-resolved event-related mental chronometry data sets

3.1 Introduction

This dissertation deals with the assessment and comparison of data-driven techniques for the analysis of complex, cognitive fMRI data sets.
In the first two chapters, we clarified the strengths and shortcomings of FCM and sICA for the analysis of an fMRI data set. Several problems arose, concerning both their use and the interpretation of their results. These problems became particularly impeding when analysing complex, cognitive data sets, i.e. when dissecting the hierarchical processing stages of the applied cognitive task.

In this dissertation, two approaches based on FCM and sICA are presented to overcome these shortcomings. Firstly, we want to deduce guidelines for the setting of their parameters in the context of a complex, cognitive task. Therefore, in this chapter a detailed elaboration on the implementing algorithms, their use and their parameters, is performed. Secondly, we expect task-induced neuronal activations and BOLD responses to produce similar signal changes in spatially contiguous regions, extending over several millimetres. Therefore, in this chapter, we present an enhanced FCM algorithm that incorporates spatial information in the detection process.

The guideline in these approaches is the improvement of the detection accuracy, consistency, and robustness, as well as on validability and interpretability of the results. In order to allow the assessment of these criteria, appropriate data sets were selected. Time-resolved event-related mental chronometry fMRI data sets were acquired in the context of a newly devised visuospatial mental imagery experiment. In this chapter, we also elaborate on the experiment as well as on the characteristics of the latter data set. The findings of this chapter will be applied in two studies in chapter 4 and 5.
Since the proposed techniques apply two different views on the data, i.e. temporal or spatial, we will first clarify the method-dependent presentation of the acquired data in a corresponding data matrix.

3.2 Spatial and temporal view on fMRI data

In this section, the arrangement of the acquired data in a matrix, prior to their analysis, is expounded. Some quantities, earlier presented in chapter 2 are thereby recalled.

An fMRI experiment results in a 4-dimensional data set, consisting of 3 spatial dimensions (the acquired volume) and 1 temporal dimension (T successive acquisitions). Preliminary to the analysis by multivariate methods, the data are presented in a 2-dimensional matrix $\boldsymbol{X}$. Therefore, the spatial dimensions are rearranged, which is accomplished in a two step procedure. Firstly, one takes advantages of the fact that functional activity is restricted only to those voxels of a functional data set that lie within the cortex. The measured volume data are therefore projected on the inflated and flattened representation of the cortical surface of the brain [1]. Subsequently, this 2 dimensional matrix representation is reshaped into a 1-dimensional array of N cortex voxels. For reasons of simplicity we will refer in the remainder of this dissertation to such array as an 'image'.

The latter rearrangement results in a $T \times N$ data matrix $\boldsymbol{X}$. Each row vector $\boldsymbol{x}_t$ of this matrix holds an image, acquired at time t ($t = 1 \dots T$). Each image $\boldsymbol{x}_t$ consists of N observed voxel amplitudes $x_{t,n}$ ($n = 1 \dots N$). Since T successive observations are made, the data matrix can be interpreted as observations of a T-dimensional variable. We emphasize that in this dissertation, dealing with specific methods for the analysis of fMRI data sets, the observed data vectors are conventionally represented by *row* vectors (as opposed to usual conventions). A column vector will be indicated by the transpose operator $(.)^T$. Whenever an elaboration uses a deviating notation, such as with SVD (see section 3.3.1), the latter will be mentioned explicitly.

Finally, we mention that in this dissertation, the mean value (i.e. the mean image) of the observed data matrix is subtracted, because we are interested in signal changes. In compact notation, this yields (for reasons of brevity we will not introduce a new variable, but refer in the rest of this section to the zero mean data matrix with $\boldsymbol{X}$):

$$\boldsymbol{X} \leftarrow \boldsymbol{X} - \bar{\boldsymbol{X}} \quad , \tag{3.1}$$

with $\bar{\boldsymbol{X}}$ the mean image, i.e. the mean over the rows.

The $T \times N$ data matrix $\boldsymbol{X}$ implies an interesting duality for its interpretation and the subsequent analysis, as shown in Fig. 3.1. In the aforementioned rearrangement procedure, a spatial view was adopted. The data matrix $\boldsymbol{X}$ was considered as a time sequence of volumes, i.e. T images, consisting of N voxel values each, are observed (Fig. 3.1a). Alternatively, adopting a temporal view, $\boldsymbol{X}$ can be considered as a spatial distribution of time courses, i.e. N time courses, consisting of T time samples each, are observed (Fig. 3.1b). In order to comply with conventions used in this dissertation, the matrix shown in Fig. 3.1b will be transposed for subsequent analysis.

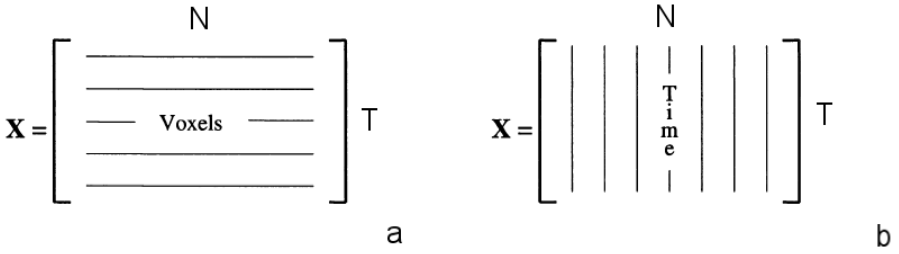

Fig. 3.1: *Two dual views on the data matrix $\boldsymbol{X}$. In the spatial view, the data matrix $\boldsymbol{X}$ is considered as a time sequence of volumes, i.e. T images consisting of N voxel values each (a). In the temporal view, $\boldsymbol{X}$ is considered as a spatial distribution of time courses, i.e. N time courses, consisting of T time samples each (b).*

Whether the temporal or spatial representation is preferred depends both on the acquisition technique, determining the dimensions of $\boldsymbol{X}$, as well as on the analysis method, imposing requirements on the characteristics of $\boldsymbol{X}$. As the data sets used in this dissertation originate from fMRI experiments only, the view on the data will depend on the adopted analysis method. In the following sections, 3.3 and 3.4, we elaborate on algorithms implementing sICA and FCM, respectively, for the analysis of the data represented by matrix $\boldsymbol{X}$. We will present extensions to the algorithms and approaches to determine the parameter settings. The focus lies on the improvement of the detection accuracy in the context of complex tasks.

3.3 Methodology of sICA-based approaches

3.3.1 Principal Components Analysis

In this dissertation, principal components analysis is applied as a preprocessing technique prior to sICA and FCM, to reduce the number of dimensions and to separate noise from signal. In this section, the singular value decomposition algorithm for the determination of the principal components is considered, as it will be applied in mathematical elaborations on ICA. For a detailed overview of alternative implementing algorithms, we refer to [2].

In the following, the spatial point of view is applied. Data vectors are represented in *columns*, which leads to deviating notations compared to the rest of the dissertation. Let the $T \times N$ data matrix $\boldsymbol{X}$ be of rank R ($R \leq \min(T,N)$). PCA is achieved by means of the singular value decomposition of $\boldsymbol{X}$:

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T \quad , \tag{3.2}$$

where $\boldsymbol{U}$ is an orthonormal $T \times R$ matrix, whose columns $\{\boldsymbol{u}_r\}$ ($r = 1 \ldots R$) are the R eigenvectors of matrix $\boldsymbol{X}$ spanning its column space. These eigenvectors have dimension $T\times$ 1, i.e. they represent the eigentime-courses of matrix $\boldsymbol{X}$. Matrix $\boldsymbol{V}^T$ is an orthonormal $R \times N$ matrix, whose rows $\{{\boldsymbol{v}_r}^T\}$ are the R eigenvectors of matrix $\boldsymbol{X}$ spanning its row space. These eigenvectors have dimension $N\times$ 1, i.e. they represent the eigenimages of matrix $\boldsymbol{X}$. The diagonal $R \times R$ matrix $\boldsymbol{D}$ contains the square root of the eigenvalues $\{d_r\}$ on the main diagonal. The data matrix can thus be written as:

$$\boldsymbol{X} = \sum_{r=1}^{R} \sqrt{d_r}\boldsymbol{u}_r{\boldsymbol{v}_r}^T \quad . \tag{3.3}$$

In the following, we consider the eigenvectors ordered such that their associated eigenvalues are ranked in descending order. Either set of eigenvectors can be used to uniquely represent the data matrix $\boldsymbol{X}$. The products $\{\boldsymbol{X}\boldsymbol{v}_r\}$ yield the set of R time courses ($T\times$ 1), for a decomposition in terms of the principal images $\{{\boldsymbol{v}_r}^T\}$. The products $\{\boldsymbol{X}^T\boldsymbol{u}_r\}$ yield the set of R spatial images ($N\times$ 1), for a decomposition in terms of the principal time courses $\{\boldsymbol{u}_r\}$.

The latter is now illustrated by a simulated example. Fig. 3.2 shows a bidimensional (i.e. $T = 2$) scatter plot of the observed (zero mean) values $\boldsymbol{x}_n$ of the image voxels. The covariance of the images is described by the covariance matrix $\boldsymbol{C}$, which is estimated from the data matrix $\boldsymbol{X}$:

$$\widehat{\boldsymbol{C}} = \frac{1}{M}\boldsymbol{X}\boldsymbol{X}^T \quad , \tag{3.4}$$

where the denominator expresses that the (maximum likelihood) estimate of the covariance matrix is calculated using M observed voxels. In the following, the equality is replaced by a proportionality, thereby omitting the denominator M:

$$\widehat{\boldsymbol{C}} \propto \boldsymbol{X}\boldsymbol{X}^T \quad . \tag{3.5}$$

Using Eq.(3.2), the covariance matrix is expressed as a function of the SVD decomposition of $\boldsymbol{X}$:

$$\widehat{\boldsymbol{C}} \propto \boldsymbol{X}\boldsymbol{X}^T = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{D}\boldsymbol{U}^T = \boldsymbol{U}\boldsymbol{D}^2\boldsymbol{U}^T \quad . \tag{3.6}$$

The PCs are found by means of the eigenvectors of $\boldsymbol{C}$ (we recall that they are ordered such that their associated eigenvalues are in decreasing order). The eigenvalues equal the variance explained by the corresponding PC and the sum of the eigenvalues equals the variance of the original observations, i.e.

$$\sum_{t=1}^{T} \sigma_t^2 = \sum_{r=1}^{R} d_t \quad , \tag{3.7}$$

with ${\sigma_t}^2$ the variance of the t-th variable (see [2] for further details). Fig. 3.2 shows the directions of the eigenvectors (PCs) and the associated eigenvalues.

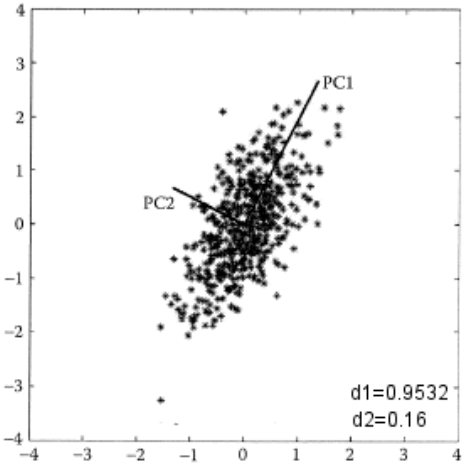

Fig. 3.2: *Bidimensional (i.e. $T = 2$) scatter plot of the observed (zero mean) values $\boldsymbol{x}_n$ of the image voxels. Principle components and associated eigenvalues of the data matrix $\boldsymbol{X}$ are shown. [3]*

3.3.2 The ICA model

3.3.2.1 Assumptions for the applicability of ICA to fMRI data sets

In a study applying ICA application to fMRI data, McKeown et al. concluded that the fMRI signals recorded during the performance of psychomotor tasks can

be decomposed into a number of independent component maps and associated component activation waveforms under the following assumptions [4, 5]:

(1) The component maps, each specified by a spatial distribution of voxel values, represent possibly overlapping multifocal brain areas related to a specific functional activation.

(2) The component map distributions are spatially independent. This is a much stronger assumption than uncorrelatedness, because independence also implies that higher-order correlations are zero.

(3) The observed fMRI signals are a linear sum of the contributions of the individual spatial processes at each voxel.

These assumptions imply that the change in the fMRI data results from the relative contributions from each of the component maps rather than of changes of the component maps themselves, i.e. the maps are fixed throughout the experiment. In order to investigate the validity of these assumptions, a number of quantities are recalled or introduced. Let $\boldsymbol{X}$ be the $T \times N$ matrix representing the spatial view on the data. Each row $\boldsymbol{x}_t$ contains an fMRI image of N voxels $x_{t,n}$ $(n = 1 \dots N)$ acquired at time t $(t = 1 \dots T)$, with T the number of scans. Let $\boldsymbol{S}$ be the $K \times N$ matrix whose rows $\boldsymbol{s}_k$ $(k = 1 \dots K)$ contain the unknown independent images $(K \leq T)$ and $\boldsymbol{A}$ the $T \times K$ linear mixing matrix whose columns ${\boldsymbol{a}_k}^T$ contain the unknown time courses of the K independent images. The matrix $\boldsymbol{A}$ is assumed to be of full rank. As illustrated in Fig. 3.3, the assumptions can be expressed rigorously by the following matrix equation:

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} \quad . \tag{3.8}$$

3.3.2.2 Conditions for the identifiability of the ICA-model

The acquired fMRI data can be modelled as a linear mixture of source-signals expressing the underlying neural activity corresponding to all stages of the task. The problem of the ICA-decomposition of fMRI time series can thus be formulated as the estimation of both matrices of the right side of Eq.(3.8). No a priori assumption is made about the mixing matrix $\boldsymbol{A}$, i.e. about the time courses corresponding to the independent images. However, in order to assure that the ICA model is identifiable, i.e. that the mixing matrix and the ICs can be estimated (up to some trivial ambiguities that will be discussed next), the following restrictions are made:

- The mixing matrix $\boldsymbol{A}$ must be square, i.e. the number T of observed mixtures is equal to the number N of independent components, as this simplifies the estimation significantly [6, 7]. In case the number of mixtures is larger, PCA can be used to reduce the dimension of the data. In general, PCA is not able to find the subspace correctly, because of the presence of noise. As a result,

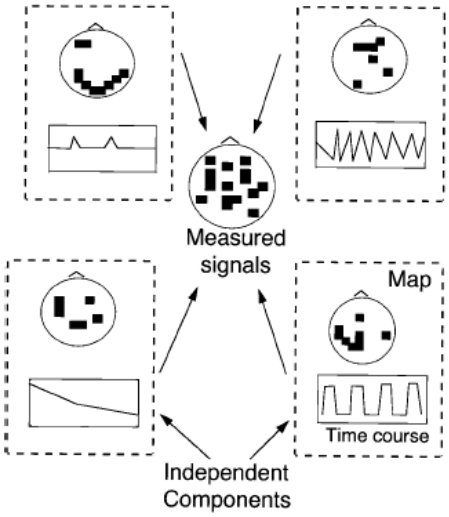

Fig. 3.3: *Schematic of fMRI data decomposed into independent components [5]. Each independent component produced by the ICA algorithm consists of a spatial distribution of voxel values ('component map'), and an associated time course of activation. The four schematic component maps (possibly partially overlapping) show voxels participating most actively in each of the four hypothetical components. The signal observed at a given voxel is modelled as the sum of the contributions of all the independent components. The amount each component contributes to the data is determined by its time course.*

some 'weak' ICs may be lost in the dimension reduction process, but PCA is still efficient in finding 'strong' ICs [8]. Moreover, reducing the dimension of the data reduces noise and prevents overlearning. The latter means that the number of parameters in a statistical model is too large compared to the number of available data points, making the estimation of the parameters difficult or even impossible. Overlearning typically produces estimates of the ICs that have a single spike or bump, and are practically zero elsewhere [9]. Finally, the unmixing matrix $\boldsymbol{W}$ (see below), which is the inverse of the square-made matrix $\boldsymbol{A}$, is well-defined, and does not require to be computed as the pseudo-inverse of $\boldsymbol{A}$. In the remainder of this section, we will assume the matrix $\boldsymbol{A}$ to be squared ($T \times T$). Consequently, T independent images are searched for, i.e. $\boldsymbol{S}$ has dimensions $T \times N$.

- The independent components must have non-Gaussian distributions. As it will be explained in section 3.3.2.6, the presence of more than one Gaussian source makes the model not undefined.

3.3.2.3 Investigating the assumptions

The assumption of statistical independence means that knowledge of the intensity values of one image $\boldsymbol{s}_k$ cannot give information about the other source images $\boldsymbol{s}_{j\neq k}$. More formally, considering the T images as random variables $\widetilde{\boldsymbol{s}}_1$, $\widetilde{\boldsymbol{s}}_2$, ..., $\widetilde{\boldsymbol{s}}_T$, the images are said to be independent if and only if the joint probability distribution function (pdf) p of these images can be expressed as the product of their marginal densities p_t:

$$p\left(\widetilde{\boldsymbol{s}}\right) = \prod_{t=1}^{T} p_t\left(\widetilde{\boldsymbol{s}}_t\right) \quad . \tag{3.9}$$

In the aforementioned study, McKeown et al. showed that, as higher-order statistics were used to enforce stricter criteria for spatial independence between maps, better estimates for the CTR components were obtained [5]. The latter suggests that spatial independence is a reasonable assumption for maps that are sparse and mostly non-overlapping. However, spatial dependence between CTR and TTR components was inferred by the changes in the TTR maps when the CTR component was removed. Also, no explicit noise model was adopted. Rather, the noise was assumed to be distributed among one or more components.

In another study, Mc Keown et al. explored the validity of the remaining assumptions, i.e. constant mixing of components throughout the brain, linear mixing, as well as the additional requirement that the number of components equals the number of time points [10]. Therefore, a representative fMRI data set was used to calculate the log-likelihood of observing each voxel's time course conditioned on the ICA model. If any of the mentioned assumptions are not valid, the ICA algorithm will be less able to separate statistically independent component maps. The estimated probability of observing the data under the null hypothesis that the ICA assumptions are valid will therefore be reduced. Although the assumptions were generally confirmed, models incorporating nonlinear mixing were studied [11].

3.3.2.4 Unmixing the data in an iterative procedure

Under the aforementioned assumptions and restrictions, the ICA decomposition of $\boldsymbol{X}$ can be defined as a linear transformation, referring to Eq.(3.8):

$$\boldsymbol{S} = \boldsymbol{W}\boldsymbol{X} \quad , \tag{3.10}$$

where the 'unmixing' matrix $\boldsymbol{W}$ is such that the independence of the target components of $\boldsymbol{S}$ is maximised. The latter is accomplished in an iterative procedure consisting of two steps:

(1) The sources $\boldsymbol{S}$ are updated as described in Eq.(3.10), using the unmixing matrix $\boldsymbol{W}$ obtained in the former iteration step.

(2) The unmixing matrix $\boldsymbol{W}$ is updated to improve the independence of the sources. The practical implementation of the applied update (or learning) rule is algorithm-dependent and will be elaborated in section 3.3.3.

Note that the linear transform $\boldsymbol{W}$ is in general unique, only up to scaling and permutation. The latter results in the following ambiguities:

- The variances (energies) of the independent components cannot be determined. Indeed, any scalar multiplier (including sign inversion) in one of the sources of $\boldsymbol{S}$ could always be cancelled by dividing the corresponding column of $\boldsymbol{A}$ by the same scalar.
- The order of the independent components cannot be determined.

3.3.2.5 Whitening the data

A whitening (or sphering) step is applied, as to transform the zero mean data matrix $\boldsymbol{X}$ into a new set of voxel values $\boldsymbol{X}_w$ that are uncorrelated and have unit variance. The covariance matrix $\widehat{\boldsymbol{C}}_w$ of the whitened data $\boldsymbol{X}_w$ thus equals the unity matrix $\boldsymbol{I}$:

$$\widehat{\boldsymbol{C}}_w = \boldsymbol{I} \quad . \tag{3.11}$$

The procedure to obtain whitened data will now be elucidated formally, i.e. for random variables $\widetilde{\boldsymbol{X}}$ with covariance matrix $\boldsymbol{C}$. The so-called whitening matrix $\boldsymbol{P}$ accomplishing the transformation

$$\widetilde{\boldsymbol{X}}_w = \boldsymbol{P}\widetilde{\boldsymbol{X}} \quad , \tag{3.12}$$

is determined by:

$$\boldsymbol{P} = \boldsymbol{C}^{-1/2} \quad . \tag{3.13}$$

Starting from the definition of the covariance matrix of $\boldsymbol{C}_w$ and taking into account Eq.(3.12) yields :

$$\boldsymbol{C}_w = \mathrm{E}\left\{\widetilde{\boldsymbol{X}}_w\widetilde{\boldsymbol{X}}_w^T\right\} = \boldsymbol{P}\,\mathrm{E}\left\{\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T\right\}\boldsymbol{P}^T \quad , \tag{3.14}$$

where E{.} is the expectation operator.
An estimate for the mixture matrix $\boldsymbol{P}$ is obtained by combining Eq.(3.13) and Eq.(3.6):

$$\widehat{\boldsymbol{P}} = \boldsymbol{U}\boldsymbol{D}^{-1}\boldsymbol{U}^T \quad . \tag{3.15}$$

As an estimate for the expected value of the covariance matrix of $\widetilde{\boldsymbol{X}}$ is obtained by Eq.(3.6), an estimate of $\boldsymbol{C}_w$ is obtained by:

$$\widehat{\boldsymbol{C}}_w = \boldsymbol{U}\boldsymbol{D}^{-1}\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{D}^2\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{D}^{-1}\boldsymbol{U}^T \quad . \tag{3.16}$$

In equations 3.15 and 3.16, the proportionality in Eq.(3.6) is replaced by an equality, as the proportionality constant is incorporated in the diagonal matrix $\boldsymbol{D}$. The latter manipulation has no impact on the validity of the proof. Indeed, it can easily be seen that, irrespective of the proportionality constant, the right hand side of equation 3.16 equals $\boldsymbol{I}$.

At the end of the whitening process, the data matrix $\boldsymbol{X}_w$ will thus have uncorrelated components with unit norm. The search for independent components will now consist in a suitable orthogonal transformation. To further clarify this concept, it is useful considering a two-dimensional example. In the left panel of Fig. 3.4, the (zero mean) raw data are plotted. In the middle panel, the whitened data are depicted. The data are now uncorrelated and have unit variance, but they are not independent. This can be seen from the figure since knowledge about the x-axis value gives information about the range of possible y-axis values. The last panel illustrates that a suitable rotation yields the independent decomposition of the original variables, as searched for by ICA.

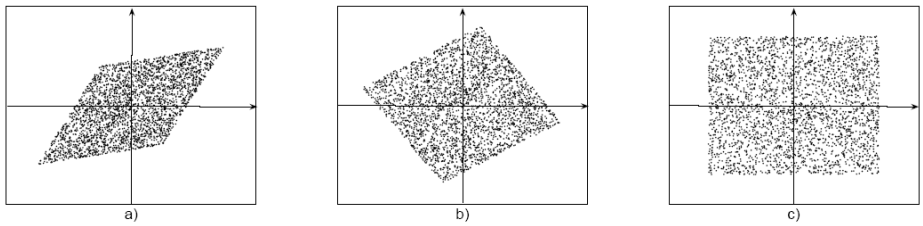

Fig. 3.4: *Example of whitening and ICA of bidimensional random data. The raw bidimensional data are shown in (a). At the end of the whitening process, the whitened data are uncorrelated and have unit variance (b). The independent components are determined by a suitable rotation (c) [6].*

Additionally, whitening reduces the computational complexity. This can be deduced as follows. Combining Eq.(3.8) and Eq.(3.12), yields:

$$\widetilde{\boldsymbol{X}}_w = \boldsymbol{P}\boldsymbol{A}\widetilde{\boldsymbol{S}} \quad . \tag{3.17}$$

Therefore, the new mixing matrix $\boldsymbol{A}_w$ can be written as:

$$\boldsymbol{A}_w = \boldsymbol{P}\boldsymbol{A} \quad . \tag{3.18}$$

The latter is orthogonal. This can be seen when reformulating the covariance matrix of $\boldsymbol{C}_w$ using Eq.(3.17) and Eq.(3.18):

$$\boldsymbol{C}_w = \mathrm{E}\left\{\widetilde{\boldsymbol{X}}_w\widetilde{\boldsymbol{X}}_w^T\right\} = \mathrm{E}\left\{(\boldsymbol{A}_w\widetilde{\boldsymbol{S}})(\boldsymbol{A}_w\widetilde{\boldsymbol{S}})^T\right\} \quad . \tag{3.19}$$

Rewriting this expression, taking into account that the matrix $\boldsymbol{S}$ is orthogonal (independent sources) yields:

$$\boldsymbol{C}_w = \boldsymbol{A}_w \mathrm{E}\left\{\widetilde{\boldsymbol{S}}\widetilde{\boldsymbol{S}}^T\right\} \boldsymbol{A}_w{}^T = \boldsymbol{A}_w \boldsymbol{A}_w{}^T \quad . \tag{3.20}$$

Since the covariance matrix $\boldsymbol{C}_w$ equals the unity matrix $\boldsymbol{I}$, the new mixing matrix is orthogonal:

$$\boldsymbol{A}_w \boldsymbol{A}_w{}^T = \boldsymbol{I} \quad . \tag{3.21}$$

Consequently, an estimate of $T(T\text{-}1)/2$ elements of the matrix $\boldsymbol{A}_w$ is needed to solve the ICA problem, instead of T^2 elements of the matrix $\boldsymbol{A}$. The complexity of the problem is thus reduced by the whitening preprocessing step.

3.3.2.6 Why Gaussian variables are forbidden

It is worth emphasizing once more that statistical independence is related to the entire pdf and therefore differs from uncorrelatedness, which is limited to second order statistics. Uncorrelatedness is a necessary but not sufficient condition for independence. For Gaussian distributions however, uncorrelatedness is equivalent to independence. This can be seen from the fact that a Gaussian distribution of an image $\boldsymbol{s}_t$ is described only by its mean and its covariance matrix, without incorporating further information in higher order statistics:

$$\boldsymbol{p}\left(\boldsymbol{s}_t\right) = \frac{1}{\left(2\pi\right)^{d/2} \left|\widehat{\boldsymbol{C}}_t\right|^{1/2}} \exp\left(-\frac{1}{2}\, \boldsymbol{s}_t\, \widehat{\boldsymbol{C}}_t^{-1/2}\, \boldsymbol{s}_t{}^T\right) \quad , \tag{3.22}$$

where d is the dimension of the multivariate distribution (in our case $d = T$) and $\left|\widehat{\boldsymbol{C}}_t\right|$ the determinant of the covariance matrix $\widehat{\boldsymbol{C}}_t$ of the (zero mean) independent source $\boldsymbol{s}_t$.

As mentioned, the mixing matrix $\boldsymbol{A}$ is whitened preliminary to the actual search for independence. As any orthogonal rotation of a multivariate *whitened* Gaussian distribution will not change its joint density distribution, there is no way to infer the mixing matrix $\boldsymbol{A}$ from the mixtures $\boldsymbol{X}$ in case of more than one Gaussian variable. Indeed, if some but not all components are Gaussian, the ICA model can estimate all the non-Gaussian components, but the Gaussian components cannot be separated from each other. Some of the estimated components will be arbitrary linear combinations of the Gaussian components. Therefore, the ICA model can still be estimated in case of just one Gaussian component.

This phenomenon can be graphically elucidated by the whitened bivariate distribution of two independent Gaussian variables, as shown in Fig. 3.5. This figure shows a rotationally symmetric density and therefore does not contain any information on the directions of the columns (containing the weights of the independent images) of the mixing matrix $\boldsymbol{A}$. Therefore, it is not possible to estimate $\boldsymbol{A}$.

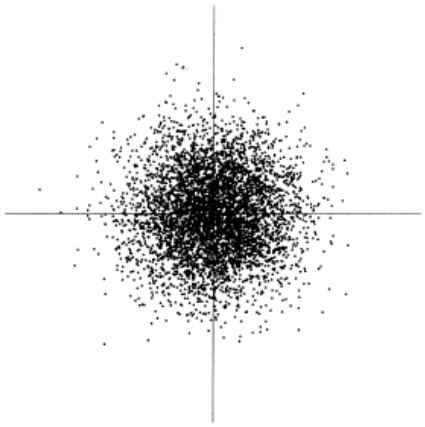

Fig. 3.5: *Gaussian variables are forbidden. The latter is illustrated by the example of the whitened distribution of a bidimensional Gaussian variable. The density is rotationally symmetric and therefore it is not possible to determine the independent components by a suitable rotation [6].*

3.3.3 Approaches and algorithms

Statistical independence of the components is accomplished when their joint probability density function factorizes, as described by Eq.(3.9). As it is rather difficult to estimate this joint probability density function, alternative approaches to measure independence are employed. Therefore, an objective function denoting independence is determined and a procedure is chosen to optimise it. Both the objective function and the optimisation procedure account for the overall performance of the ICA separation. Most objective functions proposed in literature are based on

- information theoretic criteria, such as mutual information or entropy [12], or non-Gaussianity [13].
- second-order statistics incorporating additional spatial information, such as accomplished by the lagged cross-covariances [14, 15].
- nonlinear techniques, such as nonlinear decorrelation [16, 17].

In this dissertation, we will elaborate on approaches implementing the first two criteria. Techniques implementing nonlinear decorrelation or nonlinear PCA are omitted in this dissertation, as they are considered beyond its scope. A detailed description on those techniques can be found in [6].

3.3.3.1 Information maximization

Mutual information and entropy Entropy is a basic concept of information theory, describing the uncertainty associated with a message. In the field of ICA, it will be applied to estimate the degree of independence reached in an iteration step searching for independent images $\boldsymbol{s}_t$ $(t = 1 \dots T)$. Given a set of observed images $\boldsymbol{s}_t$ constituting the matrix of all observed images $\boldsymbol{S}$, the entropy H associated with $\boldsymbol{s}_t$ is defined as (discrete-valued):

$$H\left(\boldsymbol{s}_t\right) = \sum_{j=1}^{J} p\left(\boldsymbol{s}_t = \boldsymbol{b}_j\right) \log\ p\left(\boldsymbol{s}_t = \boldsymbol{b}_j\right) \quad , \tag{3.23}$$

where $\boldsymbol{b}_j$ $(j = 1 \dots J)$ are the possible values of $\boldsymbol{s}_t$, $\mathrm{p}(\boldsymbol{s}_t = \boldsymbol{b}_j)$ is the probability that $\boldsymbol{s}_t$ is $\boldsymbol{l}_j$, and log is usually the base-2 logarithm.
The entropy $H(\boldsymbol{s}_t)$ can be interpreted as the degree of information that the observation $\boldsymbol{s}_t$ of the random variable $\widetilde{\boldsymbol{s}}_t$ gives. The more random, i.e. unpredictable the image is, the larger its entropy. For images having equal probability for all possible values, it is not possible to predict their value, which is reflected by large entropy. If there is little randomness in the image, since it takes almost always the same value, entropy is small. A remarkable property of entropy is that a Gaussian variable has the largest entropy among all the random variables having the same mean and variance.

Mutual information I can be defined as a measure of the information that some members of the set of images $\{\boldsymbol{s}_t\}$ have on the other images in the set. It is closely connected with entropy, as illustrated by the following definition:

$$I\left(\boldsymbol{s}_1, \boldsymbol{s}_2, \dots, \boldsymbol{s}_T, \right) = \sum_{t=1}^{T} H\left(\boldsymbol{s}_t\right) - H\left(\boldsymbol{S}\right) \quad . \tag{3.24}$$

The first term is related to the amount of information we get from the observation of the images separately. The second term $H(\boldsymbol{S})$ is the joint entropy of all images; it is related to the amount of information we get from the observation of all images together. The mutual information can thus be interpreted as the redundancy between the images $\boldsymbol{s}_t$ or, alternatively, as the reduction in uncertainty of a image $(\boldsymbol{s}_t)$ due to the observation of the other images $(\boldsymbol{s}_u\ u \neq t)$. In the two dimensional case, the latter can be written as

$$I\left(\boldsymbol{s}_1, \boldsymbol{s}_2\right) = H\left(\boldsymbol{s}_1\right) - H\left(\boldsymbol{s}_1 | \boldsymbol{s}_2\right) \quad . \tag{3.25}$$

If the images are statistically independent, there is no additional information about any image from the observation of any other, and the entropy of the complete set of images is the sum of the entropies of the individual images. As can be deduced from Eq.(3.24) and Eq.(3.25), in this case the mutual information equals zero. If there is some redundancy in the image set, it means that we can get some information about some image from the observation of the other images, and

the entropy of the complete set of images is lower than the sum of the individual entropies. This results in a mutual information larger than zero. The minimisation of mutual information among a set of images is thus equivalent to maximising their statistical independence.

Bell and Sejnowski: InfoMax In 1995, Bell and Sejnowski proposed the InfoMax algorithm, which determines an unmixing matrix $\boldsymbol{W}$ by attempting to maximise the joint entropy of suitably transformed component maps [12].

In order to clarify the algorithm and to demonstrate differences with PCA and correlation, a simulated experiment is described [5]. A hypothetical fMRI data set is considered as the sum of the activity of two spatially-independent processes, recorded at two separate time points. At time point $t = 1$, the subject is performing an experimental task, while time point $t = 2$ occurs during a control task condition. The two component processes, portrayed schematically in Fig. 3.6 a, are primarily active during the control and experimental task periods, respectively. Component IC2 is mostly task-related, since it is highly active at $t = 1$ and only weakly activated at $t = 2$. Component IC1 (representing either endogenous activity or machine artefact) is somewhat more active at $t = 2$ than at $t = 1$. We assume that distributions of voxel values for the two components are independent of one another, with fairly small and discrete sets of active voxels (such as those indicated for the cartoon head of Fig. 3.6 b). Here a simple reference function for detecting task-related brain areas via correlation (Fig. 3.6 c) will have the values 1 (= 'ON') at $t = 1$ and 0 (='OFF') at $t = 2$.

Figure 3.7 a shows a scatter plot of the hypothetical fMRI data. Here, for each voxel, the signal recorded at time $t = 1$ is plotted against its value at $t = 2$. The relative activations of components IC1 and IC2 (Fig. 3.6) appear in Fig. 3.7 as fixed vector directions. The assumption that component processes IC1 and IC2 are spatially independent implies that the data points will tend to be distributed uniformly along each of the vectors labelled IC1 and IC2. Note that the distribution of data values at $t = 1$ is correlated with the data distribution at $t = 2$. Thus, the marginal probability density distributions of the data in its current form are not uniform, and the data distribution cannot have maximum entropy.

The InfoMax algorithm attempts to find directions IC1 and IC2 by iteratively adjusting $\boldsymbol{W}$ so as to maximise the entropy of the resulting transformed distribution. The unmixing matrix $\boldsymbol{W}$ is initialized to the identity matrix $\boldsymbol{I}$. The iterative process consists of two steps:

(1) $\boldsymbol{S} = \boldsymbol{W}\boldsymbol{X}_w$. The unmixing matrix $\boldsymbol{W}$ obtained in the former iteration step is applied to the whitened data $\boldsymbol{X}_w$ in search for a 'more independent' unmix of the contributions of both IC-processes to the data. The linear transformation

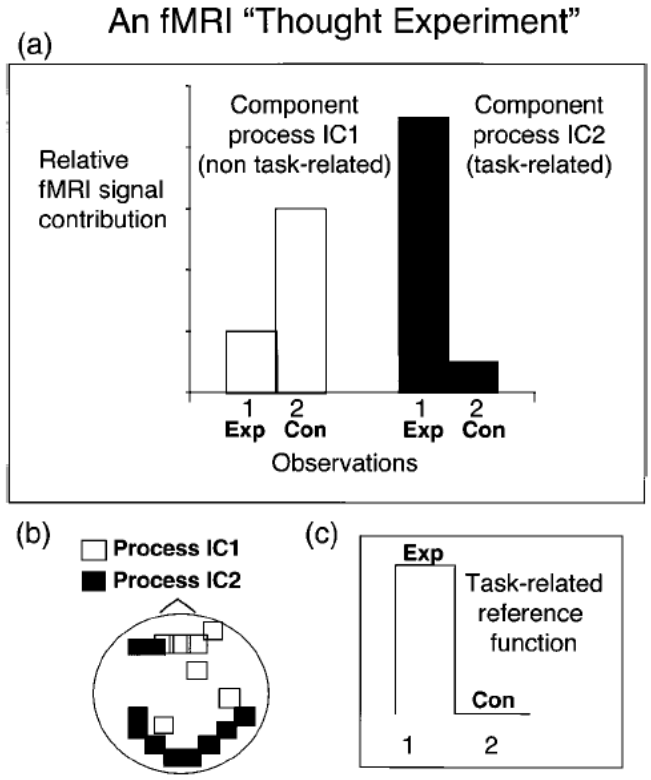

Fig. 3.6: *A simple fMRI 'thought experiment' to demonstrate differences between ICA, PCA, and correlation analysis methods (see Fig. 3.7) [5].* **a** : *A hypothetical fMRI data set is the sum of he activity of just two spatially-independent processes (IC1 and IC2) recorded at two observation times ($t = 1$, experimental; $t = 2$, control). We assume that process IC2 is more task-related, as opposed to process IC1.* **b** : *Voxels with the largest map values in the two hypothetical component distributions are active voxels of the components.* **c** : *The simplest reference function for detecting task-related activations using correlation analysis is active ($= 1$) during the experimental task and inactive ($= 0$) during the control task.*

of the scatter plot of the fMRI signal values thereby becomes more rectangular, i.e. the transformed data have larger entropy compared to the original data (Fig. 3.7 b).

(2) $\boldsymbol{Y} = g(\boldsymbol{S})$. Passing the transformed data through a nonlinearity g will more evenly spread out the data within the rectangle, producing a data distribution that has still larger entropy (Fig. 3.7 c). In order to bias the algorithm

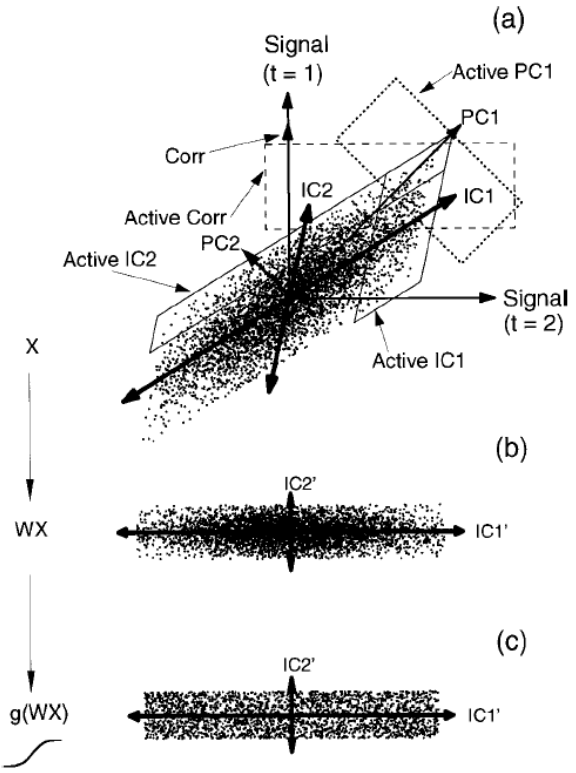

Fig. 3.7: *Analysis of the simulated experiment described in Fig. 3.6 [5].* **a** : *The scatter plot of the hypothesized fMRI signal values (at times $t = 1$ and $t = 2$) for each brain voxel contains arrows IC1 and IC2, which show the directions determined by the relative activations of the two component processes. The two parallelograms (solid borders) indicate the active voxels for each component. Active voxels by correlation analysis are those enclosed by the rectangle (dashed borders). Active voxels associated with the first principal component are those lying inside the tilted rectangle (dotted borders). ICA, PCA, and correlation analysis thus find overlapping, but typically not identical collections of active voxels. Only ICA will find the active areas of each independent component.* **b** : *The independent component directions IC1 and IC2 can be indirectly determined by finding the linear transform $\boldsymbol{W}$, which results in a rectangular distribution.* **c** : *The sigmoid transformation $g(\boldsymbol{WX})$ produces the most uniform (i.e. maximum entropy) distribution for the data shown. The ICA algorithm of Bell and Sejnowski adjusts IC1' and IC2' to maximize the entropy of the distribution. [12]*

towards finding spatially sparse component maps with relatively few highly active voxels, McKeown et al. chose to use the sigmoide [5]:

$$g(\boldsymbol{S}) = \frac{1}{1+e^{-\boldsymbol{s}}} \quad . \tag{3.26}$$

The principle behind the sigmoid function is illustrated in Fig. 3.8 for one input and one output. The steepest part of the sigmoid is aligned to the input distribution and the slope is scaled to match the variance. This flattens the output distribution and increases its entropy $H(\boldsymbol{Y})$. A gradient descent algorithm is applied, yielding an update rule for the elements of the unmixing matrix $\boldsymbol{W}$:

$$\Delta \boldsymbol{W} = \epsilon \left(\left[\boldsymbol{W}^T \right]^{-1} + g\left(\boldsymbol{W}\boldsymbol{X} \right) \boldsymbol{X}^T \right) \quad , \tag{3.27}$$

where ϵ is a learning rate that is reduced gradually during the iterative process. In practice, a user-defined threshold on a criterion expressing the change in $\boldsymbol{W}$ between two iteration steps determines when convergence is reached (e.g. the root mean square change for all elements is smaller than 10^{-6}). As soon as the latter is smaller then . The procedure is simple and smooth, but may lead to the closest local minimum instead of the global minimum.

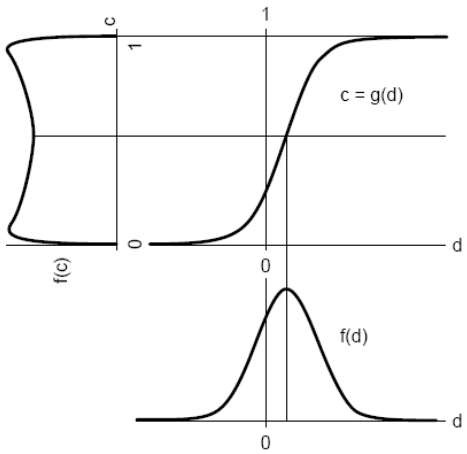

Fig. 3.8: *Maximizing entropy by a sigmoidal function [12]. The input d, with the density function $f(d)$ is passed through a nonlinear function, $c = g(d)$. The maximum amount of information the output can provide about the input is realized when the entropy of the output distribution $f(c)$ is maximized, i.e. when the output distribution is flat .*

Fig. 3.7 compares the results obtained by ICA to these obtained by PCA or correlation. Active voxels in the component maps, i.e. the voxels highlighted in Fig. 3.6b, are those that project most strongly on the vectors IC1 and IC2, as indicated

by the two parallelograms ('Active IC'). Active voxels according to correlation of the data with the assumed reference function (illustrated in Fig. 3.6c) are those whose projection on the reference function direction exceeds some threshold. The latter is indicated by the dashed rectangle ('Active Corr'). Unlike ICA, PCA finds orthogonal directions of maximum variance in the data, which has in general no specific relationship to the directions (i.e. time courses) of the independent components. Active voxels in the PC1 direction are indicated by the tilted dotted rectangle ('Active PC1'). The second component of these data has no particular relationship to either of the independent components. Results of the three methods will thus not be identical. Given the validity of the assumptions of linear summation, spatial sparsity, and statistical independence between components, ICA should more accurately determine the exact spatial extents and time courses of activations contributing to the data.

Maximum Likelihood Methods In case of sub-Gaussian distribution of sources, the presented InfoMax algorithm fails to recover the underlying sources. To overcome this drawback, Lee et al. proposed an extension of the InfoMax algorithm based on maximum likelihood (ML) estimation [18]. The likelihood function L calculates the chance that the observed images $\boldsymbol{x}_t$ of the data matrix $\boldsymbol{X}$ are generated according to a predefined model with unknown parameters. In our context, the model is the unmixing matrix $\boldsymbol{W}$. The latter approach yields

$$g_1(\boldsymbol{S}) = -2\ \tanh(\boldsymbol{S}) \quad , \tag{3.28}$$

for super-Gaussian sources and

$$g_2(\boldsymbol{S}) = \tanh(\boldsymbol{S}) - \boldsymbol{S} \quad , \tag{3.29}$$

for sub-Gaussian sources.

Due to the inversion of the matrix $\boldsymbol{W}^T$ in Eq.(3.27), needed in every iteration step, the Bell and Sejnowski algorithm converges slowly. Amari et al. pointed out that the parameter space of $\boldsymbol{W}^T$ is not Euclidean, but has a Riemannian metric structure [19, 20]. In such a case the steepest direction is given by the so-called natural gradient. The corresponding update avoids matrix inversions and thus speeds up convergence:

$$\Delta \boldsymbol{W} = \epsilon \left(\frac{\partial H(\boldsymbol{Y})}{\partial \boldsymbol{W}} \right) \boldsymbol{W}^T \boldsymbol{W} \quad . \tag{3.30}$$

3.3.3.2 Non-Gaussianity maximisation

Central limit theorem Another widely used criterion is the maximization of non-Gaussianity. The latter principle is illustrated by a simple example, considering two independent components having uniform densities. Their joint distribution is illustrated in Fig. 3.9a, in which a sample of the independent components is

plotted on the two-dimensional plane. Fig. 3.9b shows a histogram estimate of the uniform densities. These variables are then linearly mixed, and the mixtures are whitened as a preprocessing step. The joint density of the whitened mixtures, which is a rotation of the original joint density, is given in Fig. 3.10a. The densities of the two linear mixtures are estimated in Fig 3.10b. One can clearly see that the densities are closer to a Gaussian density than the densities of the independent components shown in Fig. 3.9b.

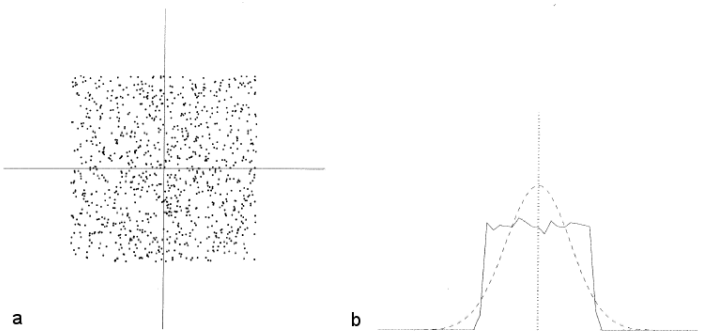

Fig. 3.9: *The joint distribution of two independent components with uniform densities (a). The estimated density of one uniform independent component, with the Gaussian density (dashed curve) given for comparison (b). [6]*

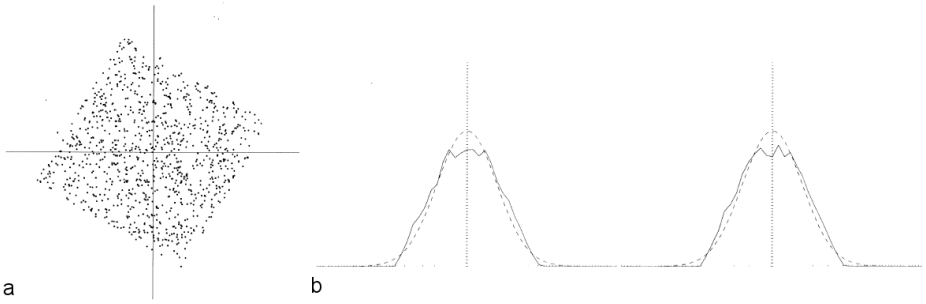

Fig. 3.10: *The joint distribution of two whitened mixtures of independent components with uniform densities (a). The marginal densities of the whitened mixtures are closer to the Gaussian density (dashed curve) than the densities of the independent components (b). [6]*

The latter principle is known as the 'central limit theorem' (CLT), stating that the linear mixing of statistically independent random signals is more Gaussian than

the original ones. It can be shown that, if the estimated ICs are constrained to have the unit variance, the principle of minimizing mutual information is equivalent to maximizing non-Gaussianity (see [6] chapter 10).

The independent components can now be retrieved by applying the CLT: an unmixing variable is maximally non-Gaussian when it equals one of the independent components. As an additional indication of problems with Gaussian components it has to be noted that if more than one component has a Gaussian probability density, it will not be possible to identify the ICs using CLT [7]. In order to estimate the degree of non-Gaussianity of an image $\boldsymbol{s}_t$, a measure must be used. A first option is to use kurtosis that is the fourth order cumulant. Kurtosis is zero for Gaussian distributions and nonzero for (almost) all non-Gaussian distributions.

$$kurt\left(\boldsymbol{s}_t\right) = \kappa_4 = \mathrm{E}\left\{\boldsymbol{s}_t^4\right\} - 3\left[\mathrm{E}\left\{\boldsymbol{s}_t^2\right\}\right]^2 \quad . \tag{3.31}$$

Hyvärinen: FastICA Kurtosis is not a robust measure with respect to noise and outliers. The value of the kurtosis may depend on only a few observations in the tails of the distribution, which may be erroneous or irrelevant observations. Assume for example that a sample of 1000 values of a random variable with zero mean and unit variance contains one value equal to 10. Then the kurtosis equals at least $10^4/1000$ - 3 = 7, which means that the single value makes kurtosis large. A more robust measure is negentropy that is defined as:

$$J\left(\boldsymbol{s}_t\right) = H\left(\boldsymbol{s}_{Gauss}\right) - H\left(\boldsymbol{s}_t\right) \quad , \tag{3.32}$$

where $H(\boldsymbol{s}_t)$ is the entropy of image $\boldsymbol{s}_t$ and $\boldsymbol{s}_{Gauss}$ is an image characterised by Gaussian distribution and the same variance matrix as $\boldsymbol{s}_t$. Entropy is not a scale-invariant property, but the difference between entropies, as in Eq.(3.32) is. Since Gaussian variables have the largest entropy among all variables with equal variance, negentropy is always larger than or equal to zero. It can thus be used as a measure of 'distance' between a random variable density and a Gaussian density with the same mean and variance. The proper estimate of the negentropy of a random variable requires the exact knowledge of the probability density of the variable, which is mostly not feasible. Therefore approximations of negentropy are used, such as:

$$J\left(\boldsymbol{s}_t\right) = k\left[\mathrm{E}\left\{G\left(\boldsymbol{s}_t\right)\right\} - \mathrm{E}\left\{G\left(\nu\right)\right\}\right]^2 \quad , \tag{3.33}$$

where k is a positive constant, ν is a Gaussian variable of zero mean and unit variance, and $\boldsymbol{s}_t$ is assumed to have zero mean and unit variance. G is a non-quadratic function, whose choice is influenced by the robustness of the estimator. The following choices of G have proven to be useful [6, 21]:

$$G_1\left(\boldsymbol{s}_t\right) = \frac{1}{a_1}\ \log\left(\cosh\left(a_1 \boldsymbol{s}_t\right)\right) \quad , \tag{3.34}$$

$$G_2\left(\boldsymbol{s}_t\right) = -\exp\left(-\frac{\boldsymbol{s}_t^2}{2}\right) \quad , \tag{3.35}$$

where $1 \leq a_1 \leq 2$ is some suitable constant, often taken equal to one.

The FastICA algorithm employs these approximations of negentropy to estimate the independent components [13]. Instead of using a gradient descent approach to find the solution, it employs a fast fixed-point iteration scheme, yielding the following learning rule:

$$\boldsymbol{W} \leftarrow \boldsymbol{X}_w\, g\left(\boldsymbol{W}^T \boldsymbol{X}_w\right) - g^{'}\left(\boldsymbol{W}^T \boldsymbol{X}_w\right) \boldsymbol{W} \quad , \tag{3.36}$$

which is followed by a normalization step:

$$\boldsymbol{W} \leftarrow \frac{\boldsymbol{W}}{\|\boldsymbol{W}\|} \quad , \tag{3.37}$$

function $g^{'}$ being the derivative of g and g the derivative of the functions suggested in Eq.(3.34) and Eq.(3.35) or the derivative corresponding to the fourth power as in kurtosis:

$$g_1\left(\boldsymbol{s}_t\right) = \tanh\left(a_1 \boldsymbol{s}_t\right) \quad , \tag{3.38}$$

$$g_2\left(\boldsymbol{s}_t\right) = \boldsymbol{s}_t\, \exp\left(\frac{-\boldsymbol{s}_t^2}{2}\right) \quad , \tag{3.39}$$

$$g_3\left(\boldsymbol{s}_t\right) = \boldsymbol{s}_t^3 \quad . \tag{3.40}$$

Given the nature of the objective function employed, it is possible to extract with FastICA one component at a time (deflation scheme) as well as all the components together (symmetric scheme) [13, 21].

- In deflation scheme, the first independent component can be estimated by calculating the direction $\boldsymbol{w}_1$ that maximizes the non-Gaussianity of $\boldsymbol{w}_1 \boldsymbol{X}$. The remaining independent components are found as the directions that maximise the non-Gaussianity of $\boldsymbol{w}_i \boldsymbol{X}$, with the constraint that $\boldsymbol{w}_i$ lies in the subspace orthogonal to the one spanned by the directions found in the previous steps. The latter constraint originates from Eq.(3.21) and the fact that the inverse ($\boldsymbol{W}$) of an orthogonal matrix ($\boldsymbol{A}_w$) is orthogonal as well.

- The symmetrical approach considers all components together and thereby prevents the propagation of errors. Actually the symmetric scheme evenly spreads the propagation error among the components.

3.3.3.3 Incorporating spatial information: lagged cross-covariances

Techniques implementing the classical ICA approach, i.e. based on information theoretic criteria, tend to neglect the intrinsic spatial structure in the data. Since it is a reasonable assumption that task-induced neuronal activations and BOLD responses produce similar signal changes in spatially contiguous regions, extending over several millimetres, data are expected to exhibit a spatial structure. Shuffling the columns of the matrix $\boldsymbol{X}$ will lose the spatial information (i.e. the relation between a voxel and its spatial location), but will not affect the results of the classical ICA techniques, as statistics are preserved. In order to use the available spatial information, techniques have been proposed to extract independent components with spatial structure. The following elaboration can also be applied to the dual case of temporal structure. As it is an objective of this dissertation to include spatial structure in the temporal FCM algorithm (see section 3.4.2) we will elaborate here on the spatial structure for reasons of comparison.

The simplest form of spatial structure is given by the linear *auto*covariance, that is the covariance between two voxels in an image:

$$cov\left(x_{t,m}, x_{t,m-\mu}\right) \quad , \tag{3.41}$$

where t refers to the t-th image and μ is the so-called spatial *lag*, a constant describing the 'distance' between the voxel m and the spatially neighbouring voxel. In addition to autocovariances, it is possible to consider *cross*covariances, i.e. covariances between neighbouring voxels of two different images t and u ($t \neq u$):

$$cov\left(x_{t,m}, x_{u,m-\mu}\right) \quad . \tag{3.42}$$

To consider all these statistics in compact form, the spatial-lagged covariance matrix is considered:

$$\boldsymbol{C}_{\mu} = \mathrm{E}\left\{\widetilde{\boldsymbol{x}}_m^T \widetilde{\boldsymbol{x}}_{m-\mu}\right\} \quad . \tag{3.43}$$

As seen in section 3.3.2.5, whitening the data is not enough to achieve independence. Therefore, higher order statistics are used to add some information. The second order statistics approaches, instead, do not take this additional information from higher-order statistics, but from the lagged covariance matrix $\boldsymbol{C}_{\mu}$. This approach starts from the consideration that if two images are independent, not only their cross-covariance will be zero, but also the lagged cross-covariances for *any* lag. This means that we are looking for a linear transformation $\boldsymbol{W}$ of data $\boldsymbol{X}$ yielding images $\boldsymbol{s}_t$ for which the following two properties hold:

$$\mathrm{E}\left\{\widetilde{s}_{i,m}\, \widetilde{s}_{j,m}\right\} = 0 \quad \forall i \neq j \quad (i,j = 1 \ldots T) \quad , \tag{3.44}$$

$$\mathrm{E}\left\{\widetilde{s}_{i,m}\, \widetilde{s}_{j,m-\mu}\right\} = 0 \quad \forall i \neq j, \forall \mu \quad (i,j = 1 \ldots T) \quad . \tag{3.45}$$

While Eq.(3.44) alone would lead to uncorrelatedness, the combination of 3.44 and 3.45 leads to independence, without using higher order statistics. Basically two techniques implementing this approach have been proposed, depending on the use of a single lag or several lags. Techniques with a single lag were elaborated by Molgedey and Schuster [14] and Tong et al., the latter proposing the 'AMUSE' algorithm [22]. Well known algorithms based on the application of several lags are 'SOBI' [23] and 'TDSEP' [15]. Compared to classical ICA techniques, second order methods have the advantage of dealing also with Gaussian sources. However, if the sources have all the same power spectra (and thus autocovariances), second order methods will fail in recovering the sources, while higher order techniques do not suffer from this limitation.

3.4 Methodology of FCM-based approaches

In this section, we will elaborate on the fuzzy clustering technique, starting from the standard partitioning-based unsupervised clustering algorithm, C-means (also called K-means). In this dissertation, clustering is performed using the Euclidean distance and adopting the temporal view on the data. Therefore, it is worth to recall the appropriate quantities and notations. Let $\boldsymbol{X}$ be the $N \times T$ matrix, whose rows contain the acquired time courses $\boldsymbol{x}_n$ $(n = 1 \dots N)$. Each time course consists of T time points $(t = 1 \dots T)$. Clustering classifies the observed time courses $\boldsymbol{x}_n$ into C clusters with centroids $\boldsymbol{v}_c$ $(c = 1 \dots C)$.

3.4.1 Approaches and algorithms

3.4.1.1 Crisp clustering: C-means

Minimizing the within-group variance The standard clustering algorithm for functional imaging is C-means [24, 25]. The clusters are a partition of the data, such that each vector $\boldsymbol{x}_n$ belongs to exactly one cluster. The goal is to find homogeneous clusters, i.e. to minimize the within-group variance ${\sigma_{\mathrm{W}}}^2$, and at the same time separable clusters, i.e. maximizing the between-group variance ${\sigma_{\mathrm{B}}}^2$. Therefore, a within-class inertia I_W and a between-class inertia I_B are defined:

$$I_{\mathrm{W}} = \frac{1}{N} \sum_{c=1}^{C} \sum_{n=1}^{N_c} d\left(\boldsymbol{x}_n, \boldsymbol{v}_c\right) \quad , \tag{3.46}$$

$$I_{\mathrm{B}} = \frac{1}{N} \sum_{c=1}^{C} N_c \, d\left(\boldsymbol{v}_c, \boldsymbol{v}_a\right) \quad , \tag{3.47}$$

where N_c is the number of voxels in cluster c, $\boldsymbol{v}_a$ is the weighted average of all cluster centres:

$$\boldsymbol{v}_{\mathrm{a}} = \sum_{c=1}^{C} \frac{N_c}{N} \boldsymbol{v}_c \quad , \tag{3.48}$$

and d is a distance measure, such as the Euclidean distance d_{E}:

$$d_{\mathrm{E}}\left(\boldsymbol{x}_n, \boldsymbol{v}_c\right) = \left\|\boldsymbol{x}_n - \boldsymbol{v}_c\right\|^2 = \sqrt{\sum_{t=1}^{T} \left(v_{c,t} - x_{n,t}\right)^2} \quad . \tag{3.49}$$

For a large class of distance measures, such as the Euclidean distance, the inertia of the cluster is minimized when the cluster centre is the average of all cluster members:

$$\boldsymbol{v}_c = \frac{1}{N_c} \sum_{n=1}^{N_c} \boldsymbol{x}_n \quad \forall c \quad . \tag{3.50}$$

Under this condition, the average cluster centre cluster $\boldsymbol{v}_a$ becomes the average of the data and I_{W} and I_{B} become the within-group variance ${\sigma_{\mathrm{W}}}^2$ and the between-group variance ${\sigma_B}^2$ respectively. It can be proven that the sum of the within-group variance and the between-group variance is constant and equal to the total data variance, regardless of the number of clusters or their compositions. Therefore minimizing I_{W} or maximizing I_{B} is equivalent. Accordingly, the within class inertia I_{W} alone provides a possible way of assessing the quality of a partition of clusters, but it cannot be used to qualitatively compare two partitions with different number of clusters. In particular, I_{W} of the optimal partition with C clusters is always higher than that of the optimal partition with $C+1$ clusters. The trivial partition of N clusters each containing one point only leads to a global minimum of I_{W}.

Iterative procedure C-means partitions the observed data into C clusters, thereby trying to minimize the objective function I_{W}. This is accomplished using an iterative procedure starting from a first hierarchical clustering step or a random cluster initialization. C-means results may depend in this initialisation step. The iterative procedure consists of two alternating steps. The first is the assignment of a voxel to the group with the smallest distance to its mean. The second is the calculation of new group means, as described by Eq.(3.50), based on the assignment. The algorithm repeats these steps until the partition does not significantly change.

3.4.1.2 Fuzzy Clustering: Fuzzy C-means

Bezdek's algorithm The similarity of the time course of voxel n to the centroid of a cluster c is expressed by the 'membership' u_{cn}. Based on this concept, Bezdek adapted the C-means objective function of Eq.(3.46) as follows:

$$I_{\mathrm{W}} = \frac{1}{N} \sum_{c=1}^{C} \sum_{n=1}^{N} {u_{cn}}^m \, d\left(\boldsymbol{x}_n, \boldsymbol{v}_c\right) \quad , \tag{3.51}$$

where $m > 1$ is the fuzziness coefficient used to 'tune out the noise' in the data. This objective function is minimised using Lagrange multipliers, under the following constraints:

$$\sum_{c=1}^{C} {u_{cn}}^{m} = 1 \quad \forall n \quad , \tag{3.52}$$

$$0 \leq u_{cn} \leq N \quad \forall c, n \quad , \tag{3.53}$$

$$0 \leq \sum_{n=1}^{N} u_{cn} \leq N \quad \forall c \quad . \tag{3.54}$$

Eq.(3.52) implies that a solution holding an empty cluster is not allowed. The latter optimisation yields an iterative procedure updating the centroids $\boldsymbol{v}_c$ and memberships u_{cn}:

$$\boldsymbol{v}_c = \frac{\sum_{n=1}^{N} {u_{cn}}^{m} \boldsymbol{x}_n}{\sum_{n=1}^{N} {u_{cn}}^{m}} \quad , \tag{3.55}$$

$$u_{cn} = \frac{1}{\sum_{k=1}^{C} \left(\frac{d(\boldsymbol{x}_n, \boldsymbol{v}_c)}{d(\boldsymbol{x}_n, \boldsymbol{v}_k)} \right)^{\frac{2}{m-2}}} \quad . \tag{3.56}$$

Bezdek suggested starting the algorithm from an initial set of membership values for the data set, expressed in matrix form as:

$$\boldsymbol{U}^0 = \left(1 - \frac{\sqrt{2}}{2} \right) \boldsymbol{U}_1 \; + \; \frac{\sqrt{2}}{2} \boldsymbol{U}_2 \quad , \tag{3.57}$$

with $\boldsymbol{U}_1$ a matrix with all elements equal to $1/C$ and $\boldsymbol{U}_2$ a matrix of randomly chosen elements. The procedure is stopped when successive iterations do not further change significantly memberships and cluster centres, yielding a convergence of the objective function I_{W}. In practice a user-defined threshold for change in I_{W} in terms of percentage (e.g. 10^{-5}) determines when convergence is reached.

Preprocessing includes the transformation of each time series into its z-score as to avoid the clustering algorithm to classify the voxels based on signal amplitude, instead of signal shape. Additionally, PCA is performed to reduce the data dimensionality as well as the noise.

3.4.2 Incorporating spatial information

Typically, in hypothesis based analysis of fMRI data spatial information is not used and activation maps are obtained considering only the temporal relation between each single voxel's time course and a specified model. Similarly, in fMRI applications of clustering, the assignment of a voxel to a specific cluster is only based on its temporal relation to the cluster centroid, and thus potential information from voxels in the spatial proximity of the examined voxel is ignored. Because task-induced neuronal activations and BOLD responses are expected to produce

similar signal changes in spatially contiguous regions, extending over several millimetres, it would be useful to use this additional information in the detection process. A few applications incorporated various forms of spatial information to improve detection accuracy in hypothesis based analysis. Descombes et al. used a 'Markov Random Field' (MRF) to perform a signal restoration which spatially smoothes the noise but at the same time preserves the signal shape [26]. Kiebel et al. accounted for the spatial smoothness of BOLD response with anatomically informed basis functions on reconstructed gray matter surfaces [27].

In this dissertation, a spatial function h_{cn} is introduced to include spatial information to the clustering of time series. The latter function expresses the probability that a voxel $\boldsymbol{x}_n$ belongs to a cluster c:

$$h_{cn} = \sum_{k \in R(\boldsymbol{x}_n)} u_{ck} \quad . \tag{3.58}$$

Function h_{cn} is obtained by adding memberships in a predefined neighbourhood R (with size N_{R}) but only takes into account those neighbouring voxels exhibiting a strong membership to the cluster c of interest. For the calculation of the updated cluster centre in Eq.(3.55), the 'temporal' (i.e. based on time course data only) membership function u_{cn} is now replaced by a spatio-temporal membership ${u_{cn}}^{\mathrm{s}}$ as defined by:

$${u_{cn}}^{\mathrm{s}} = \frac{{u_{cn}}^{p}\ {h_{cn}}^{q}}{\sum_{k=1}^{C} {u_{kn}}^{p}\ {h_{kn}}^{q}} \quad , \tag{3.59}$$

which describes a multiplicative model that expresses the membership in function of temporal (u_{cn}, Eq.(3.56)) and spatial (h_{cn}, Eq.(3.58)) components. Their relative influence is controlled by p and q, 2 exponents that vary independently. The denominator ensures each membership value for ${u_{cn}}^{\mathrm{s}}$ is in the range of 0 to 1.

3.5 Time-resolved event-related mental chronometry

A large body of research in human cognition is concerned with the segregation of mental events into their presumed hierarchical processing stages, the temporal aspect of such processing being termed mental chronometry. In this section, we illustrate that advances in single-event fMRI at high magnetic field strengths, combined with the application of a time-resolved approach, allows to extract the timing information in different neural substrates.

3.5.1 Single trial events

A single event refers to a single movement or the presentation of a single stimulus. In order to explain the need of a *single event*, we focus on the difference in consistency of HRs between simple and cognitive tasks. We recall that, for simple tasks, the lag of onset and the time course of the event-related fMRI signal are highly

reproducible, in spite of the temporal blurring [28–30]. This allows to average over trials to improve the fMRI signal's poor CNR (see section 1.6.1.1). In cognition however, several time-dependent modulations often occur, such as learning, alterations in strategy, habituation, and fatigue. Due to the latter phenomena, averaging looses unique information associated with each individual execution of the task. This constraint led to the need of studies imaging single events.

We further recall that the fMRI signal sensitivity was sufficient to detect small signals resulting from simple stimuli (see section 1.5) [31–34]. Importantly, several studies indicated a similar finding for the exploration of higher-order cognitive functions, which are typically associated with smaller responses. [35–37]. The first studies about cognition however, still performed *averaging over individual trials*, each corresponding to the applied single event [35]. Subsequent studies applied higher magnetic fields and achieved higher SNR levels. They were able to detected activation from *single trial events* without the need to average [36, 38, 39]. Temporal resolution however is larger in the latter case. For averaged trials, the latter is about 50 ms in certain cases [40]. True single trials can distinguish responses separated by 1-2 s [36, 39].

3.5.2 Mental chronometry

Mental chronometry provides information regarding the temporal aspects of brain activation, such as the onset and the duration of neuronal activity in a brain area, as well as the order of activations across different brain areas. [41]. A wide variety of studies was performed, generally averaging single trials with repetition times on the order of 30 s. The latter requirement assures that the fMRI response recovers completely to baseline prior to the next trial.

A vast amount of studies indicates that both the sequence and the duration of neuronal activity in the brain are revealed by BOLD fMRI. Excellent reviews, considering the applications of cognitive neuroimaging and mental chronometry can be found in [42] and [43], respectively. However, the application of BOLD-based fMRI in the study of mental chronometry is hampered by several factors. We will now elaborate on these factors and indicate how to deal with them in practice.

- The BOLD signal reflects a delayed haemodynamic response to neuronal activity, though the exact mechanism of the coupling between brain activity and vascular response is not well understood (see section 1.4). This sets a limit on how consistent the vascular response might be between different brain areas [30]. Therefore, the analysis results have to be interpreted carefully.
- Functional MRI signals are thought to reflect neuronal activity in the brain.

However, any discussion of brain metabolic demand and its effect on blood flow should consider the presence of a second main cell type in the brain: glial cells. Glial cells or simply glia are non-neuronal cells that, amongst others, provide support and nutrition to neurons. They are estimated to outnumber neurons by at least 10 to 1. Moreover, the time scale of their activity is generally slower than the spiking activity or neurons. However, measurements elucidating how *both* processes are reflected in the BOLD signal have yet to be done.

- When neurons fire, they typically have one of two effects on the 'receptor neuron': an excitatory effect, which increases the likelihood that the receptor neuron will fire in its turn and an inhibitory effect, which decreases the likelihood that the receptor neuron will fire in its turn. Recent reports give opposite indications concerning differentiating inhibitory from excitatory metabolic activity response. Waldvogel et al. suggested that BOLD signal changes corresponding to inhibitory activity are not likely to be seen [44], while Richter et al. suggested otherwise [38].

- Given that a single voxel in a typical fMRI experiment may contain 10^7 neurons, fMRI (as well as other techniques) is measuring local population averages. Therefore, many of the features of the individual neurones will be lost. There is no particular reason to believe that all the neurons are firing synchronously within the measured volume element. Consequently, if the neuronal code depends on the phase of the firing neurones, this information will likely be lost. This is certainly the case in some *signalling* pathways. In many cognitive tasks, neural processing lasts from a few hundred milliseconds to a few seconds [45]. Events on this time scale can be revealed with EPI, which acquires BOLD fMRI images on a time scale of tens of milliseconds. Therefore, it is suggested that fMRI is applicable to examine timing relationships between neurons when investigating *behaviour*.

3.5.3 Time-resolved fMRI

A time-resolved approach links one or more fMRI signal characteristics to a behavioural feature or presentation of a stimulus [42, 46]. A time-resolved approach is advantageous in several ways. It allows to 'time-lock' the data. This concept yields the temporal alignment of fMRI data of different trials, prior to averaging. Moreover, a time-resolved approach allows to be conclusive about the temporal sequence of neural activity between different regions from the acquired time courses. Functional MRI responses in all regions indeed are intrinsically different and therefore it is not straightforward to compare temporal characteristics between regions. This was indicated by many studies. Buckner et al., for example, observed that activation in the left prefrontal cortex language area was delayed about 1-2 s relative to activity in extrastriate areas during the performance of a word generation task [47].

In order to clarify the concept of time-resolved fMRI, a hypothetical experiment is considered [42]. A cognitive task consisting of three neural components is assumed: a visual presentation of the task (green), a cognitive process invoked by the task (red), and a motor component involved in responding to the task (blue), such as a push button response (Fig. 3.11A). Two trials are considered. The visual and motor components remain constant during the two trials. The cognitive processing component on the other hand varies in duration from trial to trial due to a different processing load. Its onset is assumed to be constant. The fMRI responses in the corresponding brain areas are shown in Fig. 3.11B for both trials. The time course of the visual component is consistent over trials. The shape of the time course related to the motor component is similar for both trials, since the motion duration for a button press is invariably the same. The onset however occurs at variable times, due to the cognitive component. The latter induces fMRI responses with the same onset time, but different duration depending on task difficulty. Figure 3.11C illustrates the correlation between behaviour, such as reaction time, with temporal characteristics of fMRI signals, such as onset time (dotted lines) and width (solid lines). Brain regions exhibiting a significant correlation are assumed to reflect activity involved in the processing strategy.

In practice, stimulus onset is often linked to the time to recognize an object [34, 36, 40], whereas stimulus width is often linked to the time to perform a mental rotation [39].

3.6 Visuospatial mental imagery

Mental imagery refers to the generation and manipulation of mental representations in the absence of sensory stimulation and is a core element in numerous cognitive processes. Many authors studied the cortical mechanisms underlying imagery and the spatial analysis in the visual, auditory, or motor domain using event-related fMRI during an appropriate task (see section 1.5.2). In this dissertation, *visuospatial* mental imagery is considered. In this section, we introduce a visuospatial mental imagery experiment and elucidate the cortical regions involved in its execution.

3.6.1 The visuospatial mental imagery experiment

In this dissertation, data originating from a visuospatial mental imagery experiment were analysed. The introduction of the experiment as well as the acquisition of the data was performed in the context of a study of Staeren et al. [48]. They analysed data resulting from this experiment with the 'Granger Causality Mapping' method, which allows to extract directed connections from fMRI data without requiring an a-priori specification of the underlying anatomical model [49]. The latter study was performed early 2005, at the University of Maastricht. It

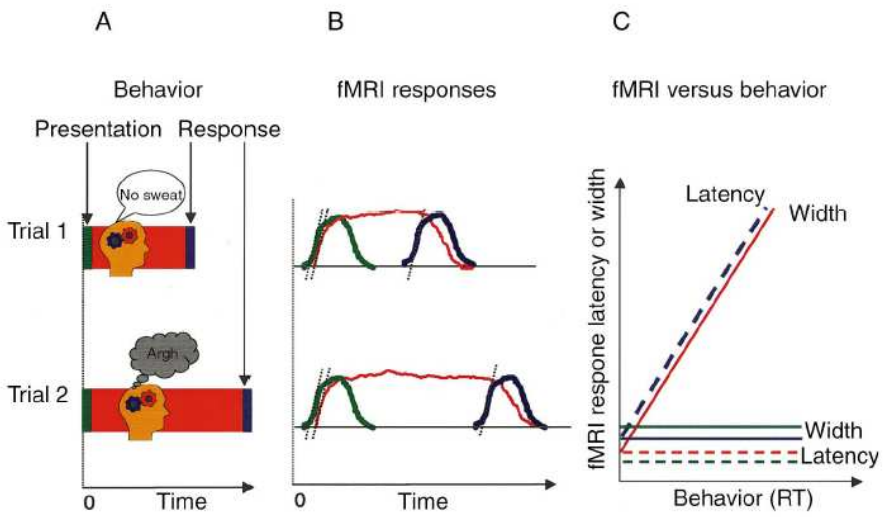

Fig. 3.11: *A hypothetical experiment, consisting of two trials, explaining time-resolved event-related fMRI [42].* **A** : *A cognitive task is assumed with three different neural components: a visual presentation of the task (green), a cognitive process invoked by the task (red), and a motor component (button press) involved in responding to the task (blue). The visual and motor components remain constant during the two trials. The cognitive processing component varies in duration from trial to trial due to a different processing load. Its onset is assumed to be constant.* **B** : *The fMRI responses in the corresponding brain areas. The shape of the time course related to the motor component is similar for both trials, since the motion duration for a button press is invariably the same. The onset however occurs at variable times depending on the processing load.* **C** : *Behaviour, such as reaction time is correlated with temporal characteristics of fMRI signals, such as onset time (dotted lines) and width (solid lines). Brain regions exhibiting a significant correlation are assumed to reflect activity involved in the processing strategy.*

used BrainVoyagerQX (Brain Innovation, Maastricht, The Netherlands), a software package for the analysis of fMRI data sets and subsequent visualization, in a research environment. The study of Staeren et al. was in the line of the research of this dissertation. The shared interest in the research topic led to contacts resulting in an intensive training course at the fMRI research unit of the Department of Psy-

chology of the University of Maastricht. This training course provided up-to-date know-how in the field of mental imagery experiments as well as experience with the BrainVoyagerQX software. Moreover, it led to a close and lasting collaboration with the fMRI research group and resulted in two submitted papers, which are the subject of chapters 4 and 5, respectively.

The experiment was specifically devised for the extraction of spatiotemporal patterns of activation as well as for the investigation of directed interactions between active cortical regions. Subjects were required to online mentally construct abstract geometric figures by juxtaposing six sequentially either visual or auditory presented stimuli. A typical geometric figure, composed of six blocks is shown in Fig. 3.12 (see 'Response'), which visually represents the design of the experiment. In our study, data referring to *auditory* instructions were analysed.
Each trial starts with a fixation cross, followed by a delay, and the sequential presentation of the six stimuli. After the construction phase, a jittered delay precedes the presentation of the target stimulus. The target stimulus is a figure composed of six blocks of which only the contours are shown by black lines. These target stimuli are either a copy or mirror image of the mentally constructed cube assembly, and are rotated by 0, 40, 80 or 120. Subjects have to indicate via button press whether the presented target stimulus is identical or mirror-reversed with respect to the constructed mental image.

This task involves a sequence of sensory, cognitive and motor processes which can take up to several seconds and can thus be studied applying fMRI mental chronometry. Staeren et al. suggested effective connectivity patterns between parietal, frontal and premotor regions [48].

3.6.2 Cortical regions involved in visuospatial mental imagery

We recall that a detailed description of the functional regions involved in the execution of the task, as well as the localization of the corresponding anatomical structures of the brain can be found in [50] and [51], respectively.

Several studies provided evidence that the analysis of visual space in perception and imagery has a common neural basis in the parietal lobes [52] and that the posterior parietal cortex (PPC) is strongly involved in the processing of spatially coded material in the imagery domain [53–56].
The functional differentiation between various cortical areas involved in visuospatial mental imagery was investigated by Trojano et al. [52] and Formisano et al.[57]. In the latter study, the cortical mechanisms underlying imagery during the mental clock task were investigated. Following results were found, as illustrated in Fig. 3.13:

- The most prominent activation related to imagery was found along the superior part of the intraparietal sulcus (IPS) and in the superior parietal lobule (SPL).

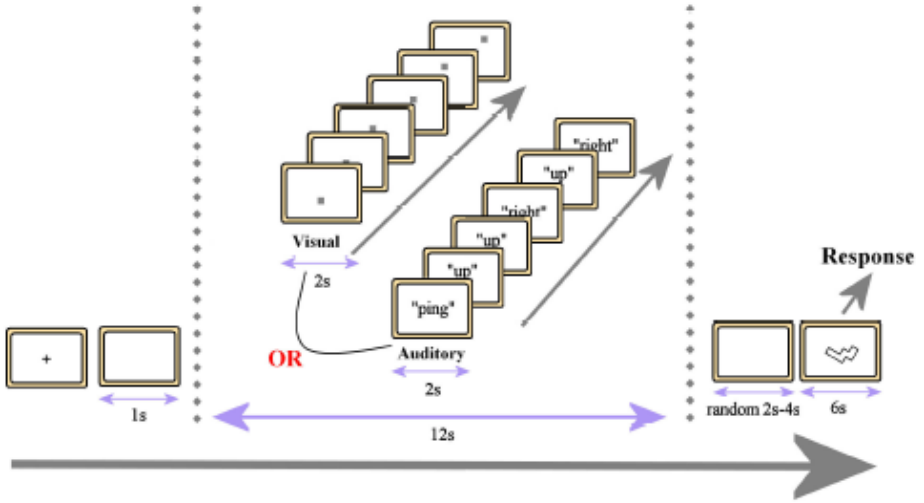

Fig. 3.12: *The visuospatial mental imagery experiment, in which a mental representation of simple two-dimensional geometric figures is created. The task involves a sequence of sensory, cognitive, and motor processes which can take up to several seconds.*

- An asymmetric sequential activation of the left and the right PPC was noticed, suggesting that these regions perform distinct functions in the imagery task.

- A part of the right PPC also participated in earlier activation, indicative of a transition from an early, more distributed, processing stage (presumably the construction of the mental representation of the angular information) to a later stage (presumably the spatial analysis) that is largely confined to the right PPC. This hypothesis that early and late clusters in the PPC participate in different stages of the task was confirmed (in the same study) by a trial-by-trial analysis of behavioural correlates of the HRs of the activated brain regions and by a subsequent study using repetitive transcranial magnetic stimulation (rTMS) [58].

3.7 Conclusions

In this dissertation, approaches and algorithms based on FCM and sICA are proposed as to enhance the analysis results of complex, cognitive data sets. We particularly aim to present guidelines for the use of both methods and for the setting of their parameters in the contect of a complex, cognitive task. Moreover,

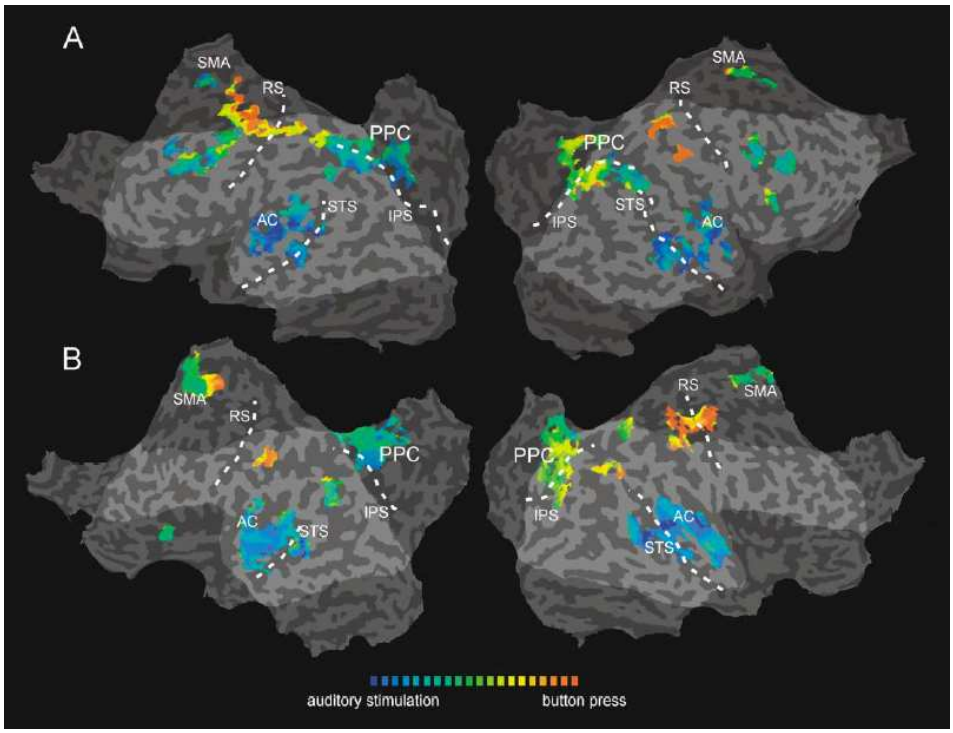

Fig. 3.13: *The sequential cortical activation from auditory perception to motor response corresponding to the mental clock task is shown [57]. Surface map are superimposed on a flattened representation of the subject's cortex in case of response with the right (A) and with the left (B) hand. The colour of significantly task-related voxels ($p < 0.001$, corrected) encodes the latency of BOLD activation following the auditory presentation of the stimulus. Blue (red) colour indicates early (late) latencies of task-related activation corresponding to the auditory stimulation (motor response). Intermediate latencies are linearly represented according to the colour bar. A different degree of lateralisation of early and late activation in the left and right posterior parietal cortex is suggested. AC= auditory cortex, IPS = intraparietal sulcus (posterior branch), PPC = posterior parietal cortex, RS = rolandic sulcus, SMA = supplementary motor area, STS = superior temporal sulcus.*

we aim at the enhancement of the inherent detection accuracy of the method, by incorporating spatial information in the detection process.

Therefore, in this chapter, we elaborately reviewed the implementing algorithms

and focused on their use and their parameter settings. Moreover, additional information of the spatial layout of the activations was incorporated in the fuzzy clustering detection process, yielding 'spatio-temporal FCM'.
In order to assess and compare these techniques and approaches, eligible data sets were acquired. Therefore, a newly devised visuospatial mental imagery experiment was introduced, yielding time-resolved event-related mental chronometry fMRI data sets.

In the remainder of this dissertation, we investigate the performance of both approaches. In chapters 4 and 5 two studies are presented, assessing and comparing these approaches and techniques in the context of the latter experiment. Chapter 4 compares FCM to sICA as to provide guidelines for the selection between both methods and the determination of parameter settings. Chapter 5 compares 'conventional FCM' to 'spatio-temporal FCM' as to deduce conditions for the improvement of the detection accuracy by adding spatial information.

Bibliography

Bibliography

[1] R Goebel and W Singer. Cortical surface-based statistical analysis of functional magnetic resonance imaging data. *Neuroimage*, 9:64, 1999.

[2] J E Jackson. *A users guide to Principal Components*. Wiley, New-York, 1991.

[3] M F Santarelli et al. *Advanced Image Processing in Magnetic Resonance Imaging*. Taylor and Francis, 2005.

[4] M J McKeown et al. Spatially independent activity patterns in functional MRI data during the stroop color-naming task. *Proc Natl Acad Sci USA*, 95 (3):803–810, Feb 1998a.

[5] M J McKeown et al. Analysis of fMRI data by blind separation into independent spatial components. *Hum Brain Mapp*, 6(3):160–188, 1998b.

[6] A Hyvärinen et al. *Independent component analysis*. John Wiley and Sons, 2001.

[7] J Eriksson and V Koivunen. Identifiability, separability, and uniqueness of linear ICA models. *IEEE Signal Processing Letters*, 11(7):601–604, 2004.

[8] J P Nadal et al. Blind source processing in the presence of weak sources. *Neural Networks*, 13(6):589–596, 2000.

[9] A Hyvärinen et al. Spikes and bumps, artefacts generated by independent component analysis with insufficient sample size. In *Proceedings International Workshop on Independent Component Analysis and Signal Separation (ICA '99)*, pages 425–429. Aussois, France, 1999.

[10] M J McKeown and T J Sejnowski. Independent component analysis of fMRI data: examining the assumptions. *Hum Brain Mapp*, 6(5-6):368–372, 1998.

[11] T W Lee et al. Blind source separation of nonlinear mixing models. In J. Principe, editor, *Proceedings of IEEE International workshop on Neural Networks for Signal Processing*, pages 406–415, Florida, 1997. New York, IEEE Inc.

[12] A J Bell and T J Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput*, 7(6):1129–1159, Nov 1995.

[13] A Hyvärinen and E Oja. A fast fixed-point algorithm for independent component analysis. *Neural Comput*, 9(7):1483–1492, 1997.

[14] L Molgedey and H G Schuster. Separation of a mixtures of independent signal using time delayed correlations . *Physical Review Letters*, 72, 1994.

[15] A Ziehe and K Muller. TDSEP - an efficient algorithm for blind source separation using time structure. pages 675–680, Proceeding of Int. Conf. on Artificial Neural Networks (ICANN'99), 1997.

[16] C Jutten and J Hérault. Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture. *Signal Proc*, 24:1–10, 1991.

[17] Cichocki A and Unbehauen R. Robust neural networks with on-line learning for blind identification and blind separation of sources. *IEEE Trans. On Circuits and Systems*, 43(11):894–906, 1996.

[18] TW Lee et al. Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and supergaussian sources. *Neural Computation*, 11(2):409–433, 1999.

[19] S Amari et al. A new learning algorithm for blind source separation. *Advances in Neural Information Processing Systems 8*, pages 757–763, 1996.

[20] S I Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):1875–1883, 1998.

[21] A Hyvärinen. Fast and robust fixed point algorithms for independent component analysis. In *IEEE Transactions on Neural Networks*, volume 10, pages 626–634, 1999.

[22] L Tong et al. Indeterminacy and identifiability of blind identification . *IEEE Trans. on Circuits and Systems*, 38, 1991.

[23] A Belouchrani et al. A blind source separation technique based on second order statistics . *IEEE Trans. on Signal Processing*, 45, 1997.

[24] J MacQueen. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1:281–297, 1967.

[25] J A Hartigan and M A Wong. Algorithm AS136. A K-means algorithm. *Appl. Stat.*, 28:100–108, 1979.

[26] X Descombes et al. Spatio-temporal fMRI analysis using Markov random fields. *IEEE Trans Med Imaging*, 17(6):1028–1039, Dec 1998.

[27] S J Kiebel et al. Anatomically informed basis funtcions . *Neuroimage*, 11: 656–667, 2000.

[28] A M Dale and R L Buckner. Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, 5:329–340, 1997.

[29] R S Menon et al. Spatial and temporal resolution of functional magnetic resonance imaging. *Biochem Cell Biol*, 76(2-3):560–71, 1998b.

[30] F M Miezin et al. Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage*, 11(6):735–59, 2000.

[31] A M Blamire et al. Dynamic mapping of the human visual cortex by high-speed magnetic resonance imaging. *Proc Natl Acad Sci USA*, 89(22):11069–11073, Nov 1992.

[32] P A Bandettini et al. Processing strategies for time-course data sets in functional MRI of the human brain. *Magn Reson Med*, 30(2):161–173, Aug 1993.

[33] P A Bandettini. *The temporal resolution of MRI. In: functional MRI*, pages 205–220. Springer-Verlag, Mauer, Germany, 1999.

[34] R L Savoy et al. Pushing the temporal resolution of fMRI: Studies of very brief visual stimuli, onset variability and asynchrony, and stimulus-correlated change in noise. In *Proceedings of the Society of Magnetic Resonance Third Scientific Meeting and Exhibition*, volume 2, pages 450–..., 1995.

[35] R L Buckner et al. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc Natl Acad Sci USA*, 93(25):14878–83, 1996.

[36] S G Kim et al. Limitations of temporal resolution in functional MRI. *Magn Reson Med*, 37(4):631–6, 1997.

[37] G McCarthy et al. Infrequent events transiently activate human prefrontal and parietal cortex as measured by functional MRI. *J Neurophysiol*, 77(3): 1630–4, 1997.

[38] W Richter et al. Sequential activity in human motor areas during a delayed cued finger movement task studied by time-resolved fMRI. *Neuroreport*, 8(5): 1257–1261, Mar 1997a.

[39] W Richter et al. Time-resolved fMRI of mental rotation. *Neuroreport*, 8(17): 3697–3702, Dec 1997b.

[40] R S Menon et al. Mental chronometry using latency-resolved functional MRI. *Proc Natl Acad Sci USA*, 95(18):10902–10907, Sep 1998a.

[41] M I Posner. *Chronometric explorations of mind*. Oxford University Press, New York, 1978.

[42] R S Menon and S G Kim. Spatial and temporal limits in cognitive neuroimaging with fMRI. *Trends Cogn Sci*, 3(6):207–216, 1999.

[43] E Formisano and R Goebel. Tracking cognitive processes with functional MRI mental chronometry. *Curr Opin Neurobiol*, 13(2):174–181, Apr 2003.

[44] D Waldvogel et al. The relative metabolic demand of inhibition and excitation. *Nature*, 406(6799):995–8, 2000.

[45] G K Aguirre et al. The variability of human, BOLD hemodynamic responses. *Neuroimage*, 8(4):360–369, Nov 1998. Clinical Trial.

[46] B R Rosen et al. Event-related functional MRI: past, present, and future. *Proc Natl Acad Sci USA*, 95(3):773–80, 1998.

[47] R L Buckner et al. Functional-anatomic study of episodic retrieval. II. Selective averaging of event-related fMRI trials to test the retrieval success hypothesis. *Neuroimage*, 7(3):163–75, 1998.

[48] N Staeren et al. Mapping directed cortical interactions during visuospatial imagery using fMRI mental chronometry and Granger causality. In *Organisation for Human Brain Mapping 2005, Toronto*. Poster 1154 W - PM.

[49] A Roebroeck et al. Mapping directed influence over the brain using Granger causality and fMRI. *Neuroimage*, 25(1):230–242, Mar 2005.

[50] J P J Pinel. *Biopsychology*. Pearson, 1999.

[51] J Talairach and P Tournoux. *Co-planar Stereotaxic Atlas of the Human Brain*. New York: Thieme Medical Publications, 1988.

[52] L Trojano et al. Matching two imagined clocks: the functional anatomy of spatial analysis in the absence of visual stimulation. *Cereb Cortex*, 10(5): 473–481, May 2000.

[53] M S Cohen et al. Changes in cortical activity during mental rotation. A mapping study using functional MRI. *Brain*, 119 (Pt 1):89–100, Feb 1996.

[54] S M Kosslyn et al. Neural systems shared by visual imagery and visual perception: a positron emission tomography study. *Neuroimage*, 6(4):320–334, Nov 1997.

[55] E Mellet et al. Functional anatomy of high-resolution visual mental imagery. *J Cogn Neurosci*, 12(1):98–109, Jan 2000.

[56] M Knauff et al. Cortical activation evoked by visual mental imagery as measured by fMRI. *Neuroreport*, 11(18):3957–3962, Dec 2000.

[57] E Formisano et al. Tracking the mind's image in the brain I: time-resolved fMRI during visuospatial mental imagery. *Neuron*, 35(1):185–194, Jul 2002.

[58] A T Sack et al. Tracking the mind's image in the brain II: transcranial magnetic stimulation reveals parietal asymmetry in visuospatial imagery. *Neuron*, 35(1):195–204, Jul 2002.

Chapter 4

Dissecting cognitive stages with time-resolved fMRI data: A comparison of Fuzzy Clustering and spatial Independent Component Analysis

The work in this chapter has been published in

A. Smolders, Federico De Martino, Noël Staeren, Paul Scheunders, Jan Sijbers, Rainer Goebel, and Elia Formisano, "Dissecting cognitive stages with time-resolved fMRI data: a comparison of fuzzy clustering and independent component analysis", Magnetic Resonance Imaging, volume, number, pages, 2007.

Abstract

In combination with cognitive tasks entailing sequences of sensory and cognitive processes, event-related acquisition schemes allow using functional MRI to examine not only the topography but also the temporal sequence of cortical activation across brain regions (time-resolved fMRI). In this study, we compared two data-driven methods - Fuzzy Clustering (FCM) and Independent Component Analysis (ICA) - in the context of time-resolved fMRI data collected during the performance of a newly devised visual imagery task. We analysed a multisubject fMRI data set

using both methods and compared their results in terms of within- and between-subject consistency and spatial and temporal correspondence of obtained maps and time courses.

Both FCM and spatial ICA allowed discriminating the contribution of distinct networks of brain regions to the main cognitive stages of the task (auditory perception, mental imagery and behavioural response), with good agreement across methods. Whereas ICA worked optimally on the original time-series, averaging with respect to the task onset (and thus introducing some a priori information on the stimulation protocol) was found to be indispensable in the case of FCM. On averaged time-series, FCM lead to a richer decomposition of the spatio-temporal patterns of activation and allowed a finer separation of the neuro-cognitive processes subserving the mental imagery task.

This study confirms the efficacy of the two examined methods in the data-driven estimation of haemodynamic responses in time-resolved fMRI studies and provides empirical guidelines to their use.

4.1 Introduction

Since its introduction in 1992 [1], functional Magnetic Resonance Imaging (fMRI) is used widely to identify the spatial layout of brain activation associated with sensory stimulations, motor actions and cognitive tasks [2]. The recent combination of event-related acquisition schemes and methods for accurate estimation of the blood oxygenation level dependent (BOLD) responses allows examining, in some cases, not only the topography but also the temporal sequence of cortical activation across brain regions [3, 4].
The achievable temporal resolution is limited to a few hundred milliseconds by the sluggishness and variability of the haemodynamic responses [5]. This approach of time-resolved fMRI is thus useful particularly in the study of complex cognition, in combination with cognitive tasks entailing sequences of cognitive processes and relatively long neural processing times. A recent study, for example, used time-resolved fMRI to address the issue of functional differentiation between the various cortical regions subserving a complex task of visuo-spatial mental imagery [6].

In the analysis of time-resolved fMRI data, the accurate estimation of task-related BOLD responses assumes a particular relevance. Complex tasks normally involve the activation of extended networks of brain regions with widely different haemodynamic responses (HRs). A region involved in the maintenance in memory of one or more items, for example, may exhibit a much more sustained neural (and haemodynamic) response than a region that transiently responds to the switching between two conditions.
Additionally, within each region, a substantial degree of experimentally-induced trial-by-trial variability is to be expected. In such cases, the analysis of HRs based on conventional model (or hypothesis)-driven methods (such as the General Linear

Model [7]) may not be optimal. Indeed, these methods assume a spatially-invariant parametric model of the HR function and this may result in a different sensitivity in one or another region of the brain. Furthermore, they do not account for the trial-by-trial variability of the responses.

A complementary approach to estimate the spatio-temporal pattern of brain activation is to use data-driven methods, such as Independent Component Analysis [8] and clustering techniques [9]. In both these approaches, a new and potentially more informative representation of the data is obtained by decomposing the original time series into a set of spatio-temporal modes, without strong a priori assumptions about the temporal profile of the effects of interest.

As applied to fMRI time series analysis, ICA attempts to separate blindly data into a set of non-Gaussian and 'spatially' statistically independent modes (independent components or ICs, see below) [8]. The ability of spatial ICA to distinguish between neurophysiologically interesting sources and noise sources has been matter of investigation in several recent publications [10, 11]. Spatial ICA has been showed to outperform Principal Component Analysis (PCA). One difficulty, however, consists in the fact that ICA does not provide any intrinsic order of the ICs. The experimenter is thus confronted with the problem of selecting and interpreting a subset of these components [6, 12, 13].

Clustering techniques separate time series into several patterns according to similarity among them. A well known member of this category is the Fuzzy Clustering Method (FCM) [14]. In the first FCM applications in neuroimaging, clustering was performed directly on the time series, using the Euclidean distance to quantify the similarity among the acquired signals [14, 15].
To overcome problems related to noise, a distance measure was introduced based on the correlation between the HRF and a stimulus function, rather than the raw time series [16]. Other studies compared FCM to alternative techniques in the field of fMRI, like correlation [17] or principal component analysis [18]. Several studies dealt with other aspects, like the cluster validity problem [19–21] or the influence of higher fields on FCM results [22]. Clustering on features extracted from the fMRI time series at each voxel was also investigated [23, 24]. Typical drawbacks of clustering approaches are the need of an a priori definition of the number of clusters (cluster validity problem) and the negative influence on the algorithm's results caused by the large amount of inactivated voxels (ill-balanced data problem). As in ICA a post hoc interpretation of the resulting clusters is required.

As illustrated in [6, 25], the new spatio-temporal representation of the data obtained by either clustering or spatial ICA has several appealing properties in the context of time-resolved fMRI. Firstly, the description of the sequence of spatial patterns of brain activation is obtained blindly thus reducing the problem of having an explicit model of the HR. Secondly, each spatio-temporal mode includes

voxels with co-varying time courses of activity and thus this representation is very helpful in highlighting the simultaneous involvement of spatially remote brain regions in the same stage of the task (functional connectivity). It is thus of interest to examine similarity and differences between the results obtained using these two different data-driven approaches. A comparison of spatial ICA and clustering has been recently described in [26]. However, this comparison was limited to a block-designed experiment with a simple visual stimulation.

In the present study, we investigate and compare the ability of FCM and ICA to estimate the spatio-temporal patterns of brain activity in the context of time-resolved fMRI measurements collected during the performance of a newly devised mental imagery task. This complex task requires the construction, maintenance, and comparison to a visual target of simple geometric figures, mentally created from auditory instructions. We analyse a multisubject fMRI data set using both methods and compare their results in terms of within- and between- subject consistency and spatial and temporal correspondence of obtained maps and time courses.

4.2 Methods

4.2.1 Fuzzy clustering

For the analysis by FCM, the temporal view is adopted, i.e. the data matrix $\boldsymbol{X}$ has dimensions $N \times T$. Fuzzy clustering attempts to partition a subset of N voxels in C 'clusters' of activation [27]. This is achieved by comparing the voxel's time courses $\boldsymbol{x}_n$ $(n = 1 \dots N)$ with each other and assigning them to representative time courses, called cluster centroids $\boldsymbol{v}_c$ $(c = 1 \dots C)$, derived during this process. Fuzziness relates to the fact that a voxel is generally not uniquely assigned to one cluster only (hard clustering), but instead, the similarity of the voxel time course to each cluster centroid is determined. This is expressed by the 'membership' u_{cn} of voxel n to cluster c. For each voxel, we have:

$$\sum_{c=1}^{C} {u_{cn}}^m = 1 \quad \forall n \quad . \tag{4.1}$$

Both centroids $\boldsymbol{v}_c$ and memberships u_{cn} are updated in an iterative procedure, elaborated by Bezdek [28] and expressed by:

$$\boldsymbol{v}_c = \frac{\sum_{n=1}^{N} {u_{cn}}^m \, \boldsymbol{x}_n}{\sum_{n=1}^{N} {u_{cn}}^m} \quad , \tag{4.2}$$

$$u_{cn} = \frac{1}{\sum_{k=1}^{C} \left(\frac{d(\boldsymbol{x}_n, \boldsymbol{v}_c)}{d(\boldsymbol{x}_n, \boldsymbol{v}_k)} \right)^{\frac{2}{m-2}}} \quad , \tag{4.3}$$

where d is a distance measure, determining the similarity between the time course of a voxel and a cluster centre, and m is the 'fuzziness coefficient', determining the

fuzziness of the procedure and used to 'tune out' the noise in the data. Theoretically, m lies between 1 (smallest fuzziness) and infinity. Its ideal value, however, is problem-dependent. Several distance measures d can be defined: the Euclidean distance d_E and the Mahalanobis distance d_M [29] are mostly used and are defined as:

$$d_E\left(\boldsymbol{x}_n, \boldsymbol{v}_c\right) = \left\|\boldsymbol{x}_n - \boldsymbol{v}_c\right\|^2 = \sqrt{\sum_{t=1}^{T}\left(v_{c,t} - x_{n,t}\right)^2} \quad , \tag{4.4}$$

$$d_M\left(\boldsymbol{x}_n, \boldsymbol{v}_c\right) = \left(\boldsymbol{x}_n - \boldsymbol{v}_c\right)\ {\boldsymbol{C}_c}^{-1}\ \left(\boldsymbol{x}_n - \boldsymbol{v}_c\right)^T \quad , \tag{4.5}$$

where $\boldsymbol{C}_c$ represents the covariance matrix of cluster c. The Mahalanobis distance takes into account the actual (elliptical) shape of the cluster, i.e. instead of treating all voxels $\boldsymbol{x}_n$ equally when calculating the distance d to the cluster centre $\boldsymbol{v}_c$, it weights the differences by the range of variability, described by $\boldsymbol{C}_c$, in the direction of the voxel. The Euclidean distance does not take into account the shape of the cluster, i.e. it assumes a spherical shape, corresponding to a covariance matrix $\boldsymbol{C}_c$ with ones on the main diagonal and zeros elsewhere.

The algorithm starts from an initial set of membership values for the data set, expressed in matrix form as:

$$\boldsymbol{U}^0 = \left(1 - \frac{\sqrt{2}}{2}\right)\boldsymbol{U}_1\ +\ \frac{\sqrt{2}}{2}\boldsymbol{U}_2 \quad , \tag{4.6}$$

with $\boldsymbol{U}_1$ a matrix with all elements equal to $1/C$ and $\boldsymbol{U}_2$ a matrix of randomly chosen elements. Next, the new cluster centres and memberships are computed using Eq.(4.2) and Eq.(4.3) . The procedure terminates when successive iterations do not further change significantly memberships and cluster centres, as calculated by Eq.(4.2) and Eq.(4.3). This procedure corresponds to the minimisation of the following objective function:

$${\sigma_W}^2 = \frac{1}{N}\sum_{c=1}^{C}\sum_{n=1}^{N}\ {u_{cn}}^m\ d\left(\boldsymbol{x}_n, \boldsymbol{v}_c\right) \quad , \tag{4.7}$$

which computes the within class variance over all clusters ${\sigma_W}^2$. In practice, a user-defined threshold for change in ${\sigma_W}^2$ determines when convergence is reached.

The a priori determination of the fuzziness coefficient and the number of clusters are research topics often encountered in literature [19]. Although several heuristics are introduced, the result is often problem-dependent. Preprocessing includes the transformation of each time series into its z-score as to avoid the clustering algorithm to classify the voxels based on signal amplitude, instead of signal shape. Finally PCA is performed to reduce data dimensionality.

4.2.2 Spatial independent component analysis

For the analysis by ICA, the spatial view is adopted, i.e. the data matrix $\boldsymbol{X}$ has dimensions $T \times N$. Each row $\boldsymbol{x}_t$ contains an fMRI image (the spatial processes) of N voxels ($n = 1 \dots N$) acquired at time t ($t = 1 \dots T$), with T the number of scans. Let $\boldsymbol{S}$ be the $T \times N$ matrix whose rows $\boldsymbol{s}_t$ contain the independent images and $\boldsymbol{A}$ the $T \times T$ mixing matrix whose columns ${\boldsymbol{a}_t}^T$ contain the time courses of the T independent images and is assumed to be of full rank. The problem of the ICA-decomposition of fMRI time series can be formulated as the estimation of both matrices of the right side of the following equation:

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} \quad , \tag{4.8}$$

under the constraint that the images $\boldsymbol{s}_t$ are (in the ideal case) spatially independent. No a priori assumption is made about the mixing matrix $\boldsymbol{A}$, i.e. about the time courses corresponding to the independent images. In this model, all the spatial components, with the possible exception of one, are assumed to be non-Gaussian. Structured (non-Gaussian) artefacts in the data (e.g. head movements, machine and physiological artefacts) are not explicitly modelled, but instead are treated as independent sources and are expected to be represented in one or more of the components. The amount of statistical dependence within a fixed number of spatial components can be quantified by means of their mutual information. Thus, the ICA decomposition of $\boldsymbol{X}$ can be defined as a linear transformation:

$$\boldsymbol{S} = \boldsymbol{W}\boldsymbol{X} \quad , \tag{4.9}$$

where the matrix $\boldsymbol{W}$ (the 'unmixing' matrix) is determined such that the mutual information of the target components $\boldsymbol{s}_t$ is minimised. Note that this definition of ICA and Eq.(4.9) imply that ICs are determined up to a permutation, a multiplicative constant and to the sign.

We estimated $\boldsymbol{S}$ using cortex based ICA (cb-ICA) [30] as implemented in BrainVoyagerQX (Brain Innovation, Maastricht, The Netherlands). Cb-ICA uses individual anatomical constraints and a fixed-point ICA algorithm (FastICA) [31] and allows an optimized analysis of cortical sources. After sphering the matrix $\boldsymbol{X}$ and reduction of the temporal dimension of the data set with PCA, the hierarchical (deflation) mode of the FastICA algorithm was used, i.e. the components were estimated one-by-one. After the decomposition, voxel values of IC spatial maps were z-transformed and colour coded according to the absolute value and sign [8]. It should be noted that the z-scores do not pertain to any significance statistic, because no comparison is made to a null hypothesis.

4.2.3 Functional MRI data

Fuzzy clustering and spatial ICA were compared in the context of an fMRI investigation of visuospatial mental imagery. During the functional measurements

(see below), subjects were asked to create a mental representation of simple two-dimensional geometric figures based on a sequence of auditory instructions sequentially building up each figure. After a jittered delay, subjects had to mentally rotate the internally constructed image, in accordance with a visually presented target figure, rotated over an angle (40, 80, or 120 degrees), and had to indicate with a button press whether these were identical or mirror-inversed. This task involves a sequence of sensory, cognitive and motor processes which can take up to several seconds and can thus be studied using fMRI mental chronometry.

Three healthy, young subjects participated to the experiment. For each subject, several successive fMRI data runs were recorded in a single session: 3 for subject 'LM', 3 for subject 'AB', and 4 for subject 'CJ'. During each run, 18 trials were acquired each trial lasting 30 seconds. Functional scans consisted of 23 transversal slices acquired on a 3 T Siemens Allegra (Siemens, Erlangen Germany) scanner using a single shot gradient-echo echo-planar imaging sequence (TE=30 ms, TR=1500 ms, matrix size = 64 × 64, voxel size = 3.5 × 3.5 × 4 mm^3, 730 images). During the same session each subject underwent a high resolution T_1 weighted anatomical scan (MDEFT sequence, voxel size = 1 × 1 × 1 mm^3, 176 slices per slab, slice thickness = 1 mm, data matrix = 224 mm × 256 mm, TR = 7.92 s, TE = 2.4 ms).

4.2.4 Preprocessing

The fMRI time series were subject to a series of preprocessing steps using BrainVoyagerQX (Brain Innovation, Maastricht, The Netherlands):

(1) Slice scan time correction was performed by resampling the time courses with sinc interpolation such that all voxels in a given volume represent the signal at the same point in time.

(2) Head movements were corrected automatically minimising the sum of squares of the voxel-wise intensity differences between each volume and the first volume of each run. Each volume was then resampled in three dimensional space according to the optimal parameters using trilinear interpolation.

(3) Temporal high pass filtering was performed to remove temporal drifts of a frequency below five cycles per run.

(4) After co-registration to the anatomical images collected in the same session functional volumes were projected into Talairach space [32].

(5) For each of the original 10 functional time series (3 subjects, 3 runs for subject 'AB' and 'LM' and 4 runs for subject 'CJ') an averaged data set was obtained considering an interval of 30 seconds around the onset of each auditory instruction (2 seconds pre stimulus onset and 28 seconds post stimulus) corresponding to 20 time points. This was done to increase the signal-to-noise ratio of consistently task-related (CTR) processes. The effect of this averaging

was evaluated for both methods (FCM and cb-ICA). In case of FCM, we found that averaging over the trials was indispensable to have a reasonable detection power of CTR processes. In case of ICA, we found that averaging over the trials deteriorates the results. Therefore, in this study, we consider FCM of averaged data and ICA of the original time-series.

4.2.5 Fuzzy clustering and spatial ICA: selection of parameters and visualization

Both FCM and sICA require setting of specific parameters (fuzziness coefficient, number of clusters, number of ICs), which can have a substantial influence on the results of the analyses and the comparisons between methods. Here we determined these settings by a preliminary analysis and inspection of one functional time series (subject 'LM', run 1). For this dataset, as for the other datasets (see section 4.3), maps obtained by FCM and spatial ICA were superimposed to 3D anatomical images and/or projected on an inflated and flattened representation of the cortical sheet of the subject's brain. This latter representation allows displaying in one picture the spatial topography of the clusters/ICs, thus providing a useful tool for their interpretation.

- To determine an appropriate value of the fuzziness coefficient m, we applied FCM for a range of values of m between 1 and 3. After visual assessment of the topography and time course of the resulting clusters, 1.25 was chosen as an acceptable value for the fuzziness coefficient, which is in close agreement with literature [19].

- The number of clusters was fixed to 13 for all the extractions. We applied PCA to the data sets under investigation and typically retained 13 of the 20 total dimensions, capturing at least 90 % of its variance/covariance. Initialising to higher or lower number of clusters yield to non optimal results. In case of lower number of clusters, some 'meaningful' clusters (see below) were not found and in case of higher number of clusters they were split into several clusters.

- In order to determine the influence of the initialization step, during which membership values and cluster centres are randomly chosen, we repeated the FCM decomposition multiple times (a random data set was analysed 100 times) and results were compared across multiple extractions. No significant difference in results was observed, indicating the robustness of the method to random initialization.

- When using cb-ICA, a PCA based reduction of dimensions was performed prior to the ICA decomposition. Based on data inspection the number of retained dimensions was fixed to 30, accounting for more than 99 % of the variance/covariance of the data.

4.2.6 Fuzzy clustering and spatial ICA: selection of clusters / components and comparison

To assess and compare the results of FCM and spatial ICA we focused on the clusters/components whose representative time-courses were clearly related to the imagery task. Clusters/components were compared, qualitatively, by visual inspections of the maps and time-courses and, quantitatively, by calculating the spatial (temporal) cross-correlation coefficients between maps (time-courses). Furthermore, for both FCM and spatial ICA we analysed the consistency of the results across runs (within subjects) and subjects.

4.3 Results and Discussion

4.3.1 FCM and spatial ICA maps and time-courses

Fig. 4.1 and 4.2 show a representative example of maps and time-courses obtained using FCM and spatial ICA. In Fig. 4.1, IC- (a) and cluster- (b) maps are colour-coded and projected on a flattened reconstruction of the subject (LM, run 1) cortex. Overlay of cluster-maps corresponds to membership values in the range of 0.5 to 1. Hence, only those voxels clearly assigned to a single cluster are shown. Overlay of IC-maps corresponds to normalized amplitudes greater than 1.8. Cluster-centre time-courses (solid line) and event-related averaged IC time courses (dash-dot line) are illustrated in Fig. 4.2.

Both FCM and spatial ICA decompositions allowed highlighting the contribution of distinct networks of areas to the sequential stages of the task, with good agreement across methods.

- According to the sequential ordering of the representative time-courses (see Fig. 4.2), a first cluster/IC (light green) included regions of the transverse temporal gyrus and of the superior temporal sulcus/gyrus bilaterally, reflecting the activation of the auditory cortex at the beginning of the trial. Interesting, in some FCM decompositions (see Tab. 4.1) this 'auditory' cluster was dissected into two clusters, whose time courses and spatial topography are compatible with a distinction between early auditory regions and language specific regions (dark green).

- Following cluster(s)/IC(s) (light and dark blue) were representative of brain activation during the performance of the visuospatial mental imagery task. Two clusters and one/two ICs (see Tab. 4.1) were found to include bilateral frontal and posterior parietal regions which have been previously involved in similar tasks of mental imagery [6]. Regions of the occipito-temporal cortex were also included in these maps. The time-courses of these two clusters/ICs suggested a sequential involvement of corresponding cortical networks in early and late stages of the imagery task.

- Finally, one or two cluster(s)/IC(s) (see Tab. 4.1) reflected the activity in the occipital (red) and in the occipito-parietal and motor regions (yellow), mostly dominated by response related to the visual presentation of the target stimulus and consequent motor response.

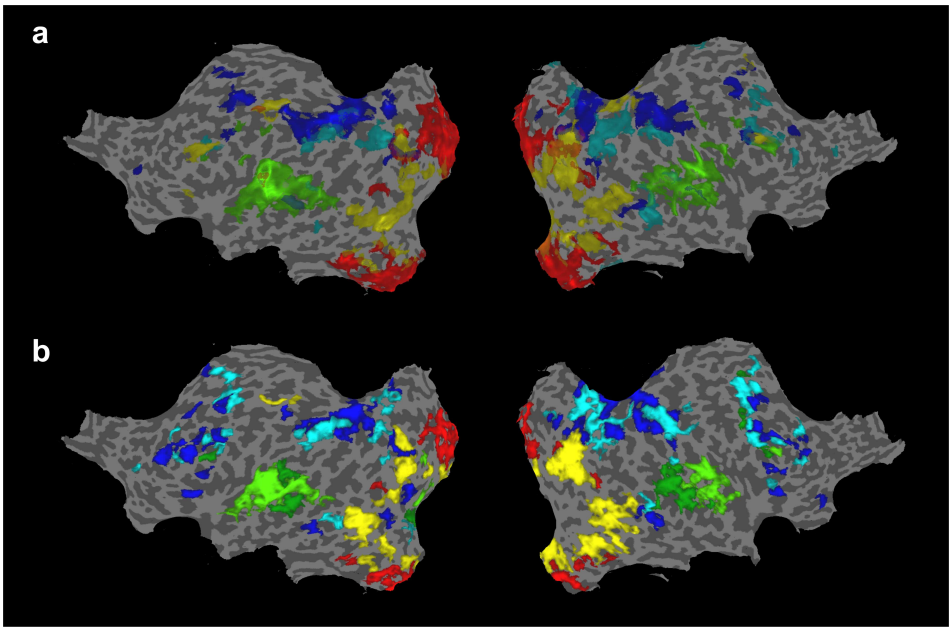

Fig. 4.1: *Projection of maps of activation regions on inflated and flattened representation of the cortical sheet of the subject's brain ('LM', run 1, left and right hemisphere) for ICA (a) and FCM (b): green = auditory; light blue = imagery 1; dark blue = imagery 2; red=visual; yellow = visual/motor. A lighter shade of a colour indicates a larger membership or IC-amplitude*

4.3.2 Comparison between methods

4.3.2.1 Within- and between subject consistency

Table 4.1 illustrates ICs and clusters found by the two methods in all the data sets analysed. In all cases, both ICA and FCM were able to dissect the original time series into ICs/clusters reflecting the three main cognitive processing stages required to perform the task. The within and between subject consistency of these results was very high. In fact, FCM and ICA decompositions typically included an 'auditory' (10/10 with both methods), an 'imagery' (10/10 with both methods) and a 'visual/motor' IC/cluster (10/10 with ICA, 9/10 with FCM). In

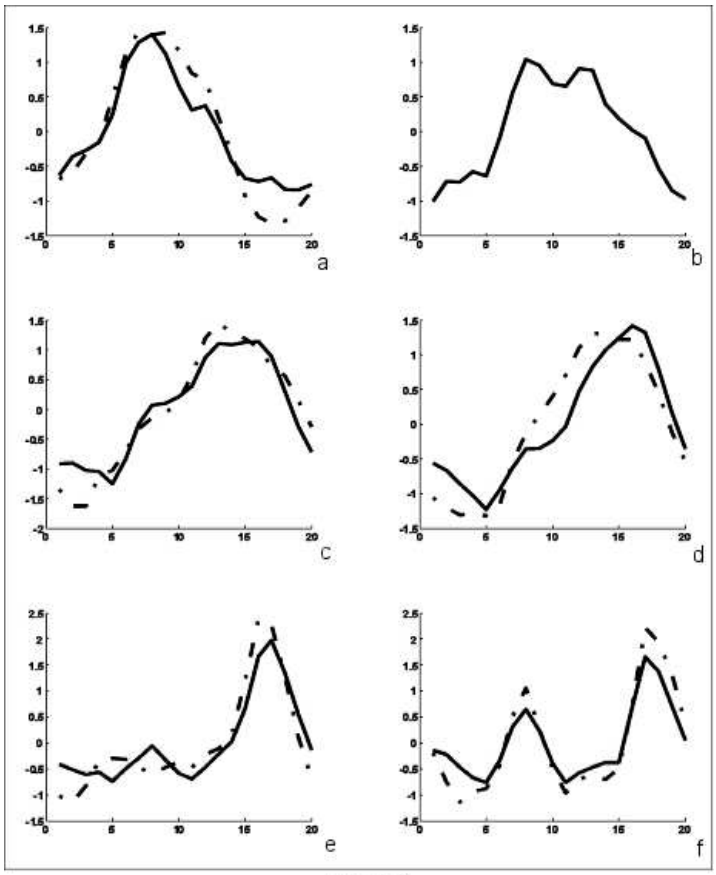

Fig. 4.2: *Time courses of corresponding processing stages found by FCM (solid line) and ICA (dash-dot line) for subject 'LM', run1.: (a) auditory 1, (b) auditory 2 (FCM only), (c) imagery 1, (d) imagery 2, (e) visual, (f) visual/motor.*

all cases, the spatial (anatomical) and temporal layout of these ICs/clusters closely resembled those described above for subject LM and were consistent with previous an expected results.

Table 4.1 also shows that, especially in the case of analysis with FCM, an even finer discrimination of neuro-cognitive stages could be reliably achieved. In particular, FCM was able to highlight in all the cases (10/10) a second cluster with fronto-parietal spatial distribution (labelled as 'imagery 2'), with a different and slightly delayed time-course with respect to imagery 1. Similarly, FCM found additional clusters presumably related to the early (auditory) and late (target related) processing stages of the tasks, respectively in five and ten cases. Note that ICA

Subject	all (10 runs)		AB (3 runs)		CJ (4 runs)		LM (3 runs)	
Component/cluster	ICA	FCM	ICA	FCM	ICA	FCM	ICA	FCM
Auditory 1	10	10	3	3	4	4	3	3
Auditory 2	1	5	0	2	1	1	0	2
Imagery 1	10	10	3	3	4	4	3	3
Imagery 2	5	10	1	3	1	4	3	3
Visual	6	10	2	3	2	4	2	3
Visual/Motor	10	9	3	3	4	3	3	3

Tab. 4.1: *Within- and between- subject consistency of components/clusters. The values in the cells indicate the number of runs in which a specific component/cluster was found.*

decompositions of the same datasets not always produced ICs that corresponded to these additional clusters. The high within- and between- subject consistency together with their anatomical layout and temporal profile suggest that this further separation is neurophysiologically meaningful and it is not due to an artificial splitting of a cluster.

4.3.2.2 Spatial and temporal correspondence between ICA and FCM decompositions

To examine the spatial and temporal similarity between ICs and clusters corresponding to the same cognitive stage we calculated, for the three main components in each data set, the spatial and temporal correlation coefficient of their maps and time courses. The temporal correlation coefficient was calculated between the time course of a cluster and the averaged time course of the corresponding IC. The spatial correlation coefficient was calculated without imposing any threshold on cluster membership or IC amplitude of spatial maps.

Figure 4.3 illustrates a summary of obtained results. Each box-plot has lines at the lower quartile, median, and upper quartile values of computed correlations. The whiskers show the extent of the rest of the data. Statistical significance threshold ($p<0.05$) for temporal correlation is 0.3783 ($T = 20$) and for spatial correlation is 0.0095 ($N\approx30.000$).
In all cases, obtained values of correlations are above this threshold, denoting a good correspondence between the maps or time courses obtained with ICA and FCM. It is noticeable that correspondence between the maps is highest in the case of the 'auditory' IC/cluster. In the case of 'imagery' and 'visual/motor' (i.e. 'target' related) ICs/clusters obtained median values of correlation are lower and a larger variability is present. This is mainly due to the fact that selected ICs include in some cases (5 for 'imagery' and 4 for 'visual/motor') spatial regions which are separated in two distinct clusters in the FCM analysis.

The previous analysis of spatial correspondence between ICs and clusters was

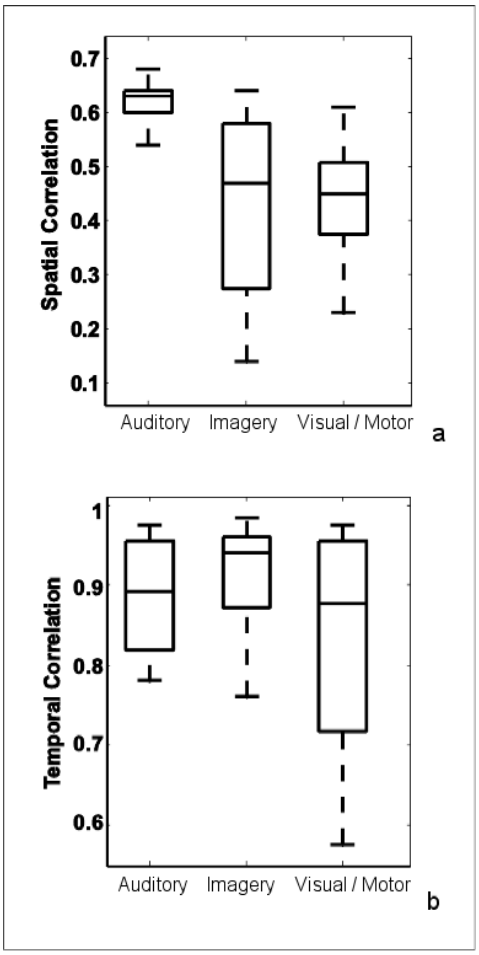

Fig. 4.3: *Boxplot of between-methods (FCM and ICA) spatial (a) and temporal (b) correlation coefficient. Each box-plot has lines at the lower quartile, median, and upper quartile values of computed correlations. The whiskers show the extent of the rest of the data.*

performed with unthresholded maps. Normally, however, interpretation of results is done on thresholded maps. It is thus of interest to examine how the correspondence between the peaks of ICA and FCM maps changes if cluster membership and IC-amplitude thresholds are varied. Figure 4.4 shows a surface obtained by calculating the spatial cross-correlation between ICA and FCM maps after they have been thresholded at varying threshold values (i.e. values in the maps below the threshold are set to 0).

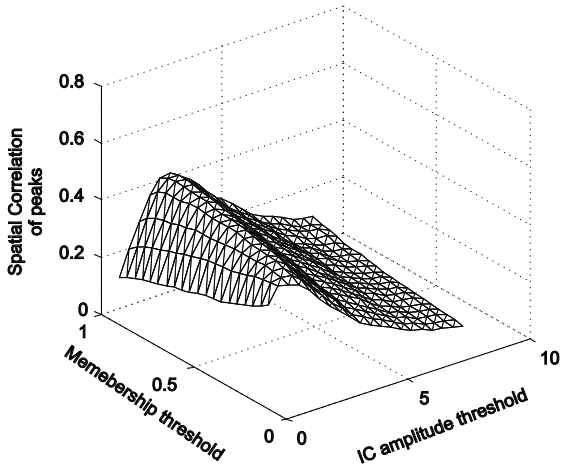

Figure 4

Fig. 4.4: *Influence of thresholds (cluster membership and IC amplitude) on spatial overlap between maps. Example for subject 'LM', run 1, spatial correlation between thresholded maps of the auditory cluster/IC found by FCM and ICA.*

It can be seen that the overlap of peaks is rather insensitive to changes in the cluster membership threshold but it can be affected to a certain extent by changes in IC-amplitude threshold. Importantly, there is a relatively large range of IC-amplitudes (1.5 - 3.5) within which there is an acceptable and constant overlap between the peaks of the two maps. Similar results were found for the other clusters.

4.4 Conclusions

In this study, we empirically compared two data-driven methods - Fuzzy Clustering and spatial ICA - in the context of time-resolved fMRI measurements. Both FCM and spatial ICA decompositions allowed highlighting the contribution of distinct networks of brain regions to the sequential stages of a visual imagery task (auditory perception, mental imagery and behavioural response). We evaluated the maps and time courses resulting from spatial ICA and FCM in terms of their within- and between- subject consistency, which resulted to be very high in both cases. Furthermore, the calculation of spatial and temporal correlations of IC/cluster maps and time courses showed a good agreement between the results obtained

with these two methods, suggesting that the partitions of the time series obtained using spatial ICA or FCM are comparable, even though they are based on substantially different principles (spatial independence vs similarity of time courses).

The analysis of our data sets also highlights a relevant distinction between the two methods and of the underlying principles. On the original (i.e. non-averaged) time series, interesting spatio-temporal patterns of activation were correctly detected by spatial ICA but not by FCM. Conversely, when time series were averaged with respect to task onset, analyses with FCM but not spatial ICA benefited from the increase in functional contrast-to-noise of task related process. On averaged time series, FCM lead to a richer decomposition of the spatio-temporal patterns of activation and allowed a finer separation of the neuro-cognitive processes subserving the mental imagery task. In particular, processes that were united in ICA in a single component were separated into distinct and presumably meaningful clusters by FCM. Note, however, that averaging of the time series implies introducing a priori information on the stimulation paradigm (but not on the shape of the haemodynamic response) and thus the subsequent analysis cannot be considered 'blind'. Also, averaging prevents examining the variability of single-trial responses and processes which are not task-related (e.g. activity in the 'default mode' network).

In sum, these results and considerations suggest complementary characteristics of FCM and spatial ICA decompositions, with an advantage of FCM in the analysis of task-related responses and averaged time series and an advantage of spatial ICA in the case the variability of single trial responses and non-task related processes are relevant.

Bibliography

Bibliography

[1] S Ogawa et al. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci USA*, 89(13):5951–5955, Jul 1992.

[2] F Di Salle et al. Exploring brain function with magnetic resonance imaging. *Eur J Radiol*, 30(2):84–94, May 1999.

[3] R S Menon et al. Mental chronometry using latency-resolved functional MRI. *Proc Natl Acad Sci USA*, 95(18):10902–10907, Sep 1998.

[4] E Formisano and R Goebel. Tracking cognitive processes with functional MRI mental chronometry. *Curr Opin Neurobiol*, 13(2):174–181, Apr 2003.

[5] G K Aguirre et al. The variability of human, BOLD hemodynamic responses. *Neuroimage*, 8(4):360–369, Nov 1998.

[6] E Formisano et al. Tracking the mind's image in the brain I: time-resolved fMRI during visuospatial mental imagery. *Neuron*, 35(1):185–194, Jul 2002.

[7] K J Friston et al. Analysis of functional (MRI) time-series. *Human Brain Mapping*, 1:153–171, 1994.

[8] M J McKeown et al. Analysis of fMRI data by blind separation into independent spatial components. *Hum Brain Mapp*, 6(3):160–188, 1998.

[9] C Goutte et al. On clustering fMRI time series. *Neuroimage*, 9(3):298–310, Mar 1999.

[10] C G Thomas et al. Noise reduction in BOLD-based fMRI using component analysis. *Neuroimage*, 17:1521–1537, 2002.

[11] M J McKeown et al. Independent component analysis of functional MRI: what is signal and what is noise? *Curr Opin Neurobiol*, 13(5):620–629, Oct 2003.

[12] J Himberg et al. Validating the independent components of neuroimaging time series via clustering and visualization . *Neuroimage*, 22(3):1214–1222, 2004.

[13] F De Martino et al. Classification of fMRI independent components using IC-fingerprints and support vector machine classifiers. *Neuroimage*, 34(1): 177–194, Jan 2007.

[14] R Baumgartner et al. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part I: reproducibility. *J Magn Reson Imaging*, 7 (6):1094–1101, Nov 1997.

[15] E Moser et al. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part II: quantification. *J Magn Reson Imaging*, 7(6):1102–1108, Nov 1997.

[16] X Golay et al. A new correlation-based fuzzy logic clustering algorithm for fMRI. *Magn Reson Med*, 40(2):249–260, Aug 1998.

[17] R Baumgartner et al. Quantification in functional magnetic resonance imaging: fuzzy clustering vs. correlation analysis. *Magn Reson Imaging*, 16(2): 115–125, 1998.

[18] R Baumgartner et al. Comparison of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis. *Magn Reson Imaging*, 18(1):89–94, Jan 2000.

[19] M J Fadili et al. On the number of clusters and the fuzziness index for unsupervised FCA application to BOLD fMRI time series. *Med Image Anal*, 5(1):55–67, 2001.

[20] U Moller et al. How to avoid spurious cluster validation? A methodological investigation on simulated and fMRI data. *Neuroimage*, 17(1):431–446, Sep 2002.

[21] W F Auffermann et al. Cluster significance testing using the bootstrap. *Neuroimage*, 17(2):583–591, Oct 2002.

[22] C Windischberger et al. Fuzzy cluster analysis of high-field functional MRI data. *Artif Intell Med*, 29(3):203–223, Nov 2003.

[24] B Thirion and O Faugeras. Feature characterization in fMRI data: the Information Bottleneck approach. *Med Image Anal*, 8(4):403–419, Dec 2004.

[23] C Goutte et al. Feature-space clustering for fMRI meta-analysis. *Hum Brain Mapp*, 13(3):165–183, Jul 2001.

[25] W Richter et al. Motor area activity during mental rotation studied by time-resolved single-trial fMRI. *J Cogn Neurosci*, 12(2):310–320, Mar 2000.

[26] A Meyer-Baese et al. Comparison of two exploratory data analysis methods for fMRI: unsupervised clustering versus independent component analysis. *IEEE Trans Inf Technol Biomed*, 8(3):387–398, Sep 2004.

[27] L A Zadeh. *Fuzzy Sets and their Application to Pattern Classification and Clustering Analysis*. Academic Press, New-York, 1977.

[28] J C Bezdek. FCM: the fuzzy C-means algorithm. *Computers and Geosciences*, 10:191–203, 1984.

[29] P C Mahalanobis. On generalized distance in statistics. In *Proceedings National Institute of Sciences India*, volume 12, pages 49–55, 1936.

[30] E Formisano et al. Cortex-based independent component analysis of fMRI time series. *Magn Reson Imaging*, 22(10):1493–1504, Dec 2004.

[31] A Hyvärinen and E Oja. A fast fixed-point algorithm for independent component analysis. *Neural Comput*, 9(7):1483–1492, 1997.

[32] J Talairach and P Tournoux. *Co-planar Stereotaxic Atlas of the Human Brain*. New York: Thieme Medical Publications, 1988.

Chapter 5

Spatio-temporal fuzzy clustering of fMRI time series

The work in this chapter has been submitted to *A. Smolders, Federico De Martino, Noël Staeren, Paul Scheunders, Jan Sijbers, Rainer Goebel, and Elia Formisano, "Spatio-temporal fuzzy clustering of fMRI time series", Neuroimage, 2007.*

Abstract

We introduce a fuzzy clustering algorithm, specifically tailored to the analysis of fMRI data sets. In contrast to previous approaches, our algorithm clusters fMRI time series based on both spatial and temporal information (spatio-temporal clustering). The probability that a voxel belongs to a cluster is weighted by a spatial function, which takes into account the neighbourhood relationships between voxels. The proposed approach is described and compared to conventional FCM on realistic simulated fMRI data sets. We use Receiver Operating Characteristics (ROC) and correlation analyses for a quantitative comparison of results. This evaluation indicates that spatio-temporal FCM performs significantly better than conventional FCM, especially in the case of low functional contrast-to-noise ratio. Also, it shows that accounting for the neighbourhood relationships between voxels during clustering is preferable to spatially smoothing of the original data. Finally, we apply spatio-temporal clustering to real data from a mental imagery experiment. Clusters obtained with the proposed method allow dissecting the

brain activation in distinct spatio-temporal patterns that reflect the various cognitive stages involved in the imagery task. Compared to conventional FCM and in agreement with the results of our simulations, these clusters present a more plausible topological structure and include additional regions exhibiting weak activation.

5.1 Introduction

In the analysis of BOLD fMRI data, estimation of spatial layout of task-related brain is conventionally based on model (or hypothesis)-driven methods, such as the General Linear Model [1]. This category of methods assumes a spatially-invariant model of the HR function. For the analysis of complex tasks however, this assumption may not be optimal. Indeed, these tasks normally involve the activation of extended networks of brain regions with widely different HRs. Additionally, within each region, a substantial degree of experimentally-induced trial-by-trial variability is to be expected [2, 3].

A complementary approach to estimate the spatio-temporal pattern of brain activation is then to use data-driven methods, such as Independent Component Analysis [4] and clustering techniques [5]. In both these approaches, a representation of the data is obtained by decomposing the original time series into a set of spatio-temporal modes, without strong a priori assumptions about the temporal profile of the effects of interest.

Clustering techniques separate time series into several patterns according to similarity among them. A well known member of this category is the Fuzzy Clustering Method (FCM) [6] that expresses the similarity of a time series to each reference pattern by a so-called membership to that pattern. FCM applications in neuroimaging typically perform clustering directly on the time series, using the Euclidean distance to quantify the similarity among the acquired signals [7–9].

Other similarity measures were introduced. To overcome problems related to noise, a distance measure was introduced based on the correlation between the time courses of the voxel and the cluster centre [10]. Alternatively the Mahalanobis distance can be used, which takes into account the elliptical shape of the clusters, whereas the Euclidean distance assumes a spherical shape [5, 11].

Other studies compared FCM to alternative techniques in the field of fMRI, like correlation [12], principal component analysis [13], neural networks [14], split-merge and region growing techniques [15], independent component analysis [16], and C-means and SOM [17]. A typical drawback of the clustering approach is the cluster validity problem.

The optimal number of clusters needs to be determined a priori [18–20] and the resulting clusters need to be interpreted post hoc [21–23].

Preprocessing often also includes techniques to solve the ill-balanced data problem, i.e. the negative influence on the algorithm's results caused by the large amount of inactivated voxels [24]. Windischberger et al. investigated the influence of higher fields on FCM results [25].

Several authors applied clustering to features extracted from the time series, instead of clustering on the raw time series. Goutte et al. investigated the use of two simple features, namely the activation strength and the response delay of the cross-correlation function between the time series and the excitation paradigm, thereby combining model-and data-driven techniques [26]. In the same study, the authors performed a meta-analysis in a seven-dimensional feature space, using different standard single-voxel analysis techniques, like Student t-test (between rest and activation) and correlation with the paradigm. Simon et al. applied clustering to data from six different task-control pairs (6 Student t-values) collected for each subject in the study [27]. Jahanian et al presented a feature space based on multiscale decompositions obtained by scalar wavelet and multiwavelet transforms [28].

Typically, in model based analysis of fMRI data spatial information is not used and activation maps are obtained considering only the temporal relation between each single voxel's time course and a specified model. Similarly, in fMRI applications of clustering the assignment of a voxel to a specific cluster is only based on its temporal relation to the cluster centroid, and thus potential information from voxels in the spatial proximity of the examined voxel is ignored.

Because task-induced neuronal activations and BOLD responses are expected to produce similar signal changes in spatially contiguous regions, extending over several millimetres, it would be useful to use this additional information in the detection process. A few applications incorporated various forms of spatial information for improving detection accuracy in model based analysis. Descombes et al. used a Markov Random Field (MRF) to perform a signal restoration which spatially smooths the noise but at the same time preserves the signal shape [29]. Kiebel et al. [30] accounted for the spatial smoothness of BOLD response with anatomically informed basis functions on reconstructed grey matter surfaces.

In the present paper, we propose an approach for adding spatial information to the 'temporal' clustering of fMRI time series by weighting cluster memberships in a 3-dimensional (3D) neighbourhood of the voxel. The probability that a voxel belongs to a cluster is expressed by a spatial function, which takes into account the neighbourhood relationships between voxels. A neighbouring voxel will have an influence on the assignment to a cluster for the voxel being investigated which depends on its membership value to this same cluster.

The model can be seen as an extension to 4-dimensional (4D) data sets of cluster-

ing algorithms used for the segmentation of images, in which spatial information is accounted for in the clustering objective function [31], membership [32], or distance measure [33].

It is worth emphasizing that the proposed method exploits spatial information during clustering as opposed to spatial smoothing in which a Gaussian filter is applied preliminarily to the measured data. Our method also differs conceptually from previous approaches which incorporate spatial information preliminary to an FCM analysis [34]. Indeed, in our approach, a neighbouring voxel which is not clearly assigned to the same cluster as the current voxel will have a limited influence on the method's outcome.

In this paper, we elaborate on spatio-temporal FCM, presenting an FCM algorithm tailored to the 4D nature of fMRI data sets, compare our approach to conventional FCM and to spatial smoothing using a realistic simulated data set, and illustrate the method's increased sensitivity in regions exhibiting low CNR. Finally, we apply our findings to analyse real data originating from a complex mental imagery experiment.

5.2 Methods and materials

In the following, the temporal view is adopted, i.e. the data matrix $\boldsymbol{X}$ has dimensions $N \times T$.

5.2.1 Conventional clustering

Fuzzy clustering attempts to partition a subset of N voxels in C 'clusters' of activation [6]. Therefore a specific set of features is derived from the voxel's time course $\boldsymbol{x}_n$ $(n = 1 \dots N)$. The resulting sets of values for all voxels are then compared and assigned to representative sets, called cluster centroids $\boldsymbol{v}_c$ $(c = 1 \dots C)$, derived during this process. Typically, as is the case in this study, features used are the voxel's time courses $\boldsymbol{x}_n$ themselves. Clustering is thus performed in the time domain. Fuzziness relates to the fact that a voxel is generally not uniquely assigned to one cluster only (hard clustering), but instead, the similarity of the voxel time course to each cluster centroid is determined. This is expressed by the 'membership' u_{cn} of voxel n to cluster c. For each voxel, we have:

$$\sum_{c=1}^{C} {u_{cn}}^m = 1 \quad \forall n \quad . \tag{5.1}$$

Both centroids $\boldsymbol{v}_c$ and memberships u_{cn} are updated in an iterative procedure, elaborated by Bezdek [35] and expressed by:

$$\boldsymbol{v}_c = \frac{\sum_{n=1}^{N} {u_{cn}}^m \, \boldsymbol{x}_n}{\sum_{n=1}^{N} {u_{cn}}^m} \quad , \tag{5.2}$$

$$u_{cn} = \frac{1}{\sum_{k=1}^{C} \left(\frac{d(\boldsymbol{x}_n, \boldsymbol{v}_c)}{d(\boldsymbol{x}_n, \boldsymbol{v}_k)} \right)^{\frac{2}{m-2}}} \quad , \tag{5.3}$$

where d is a distance measure, determining the similarity between the time course of a voxel and a cluster centre, and m is the 'fuzziness coefficient', determining the fuzziness of the procedure and used to 'tune out' the noise in the data. Theoretically, m lies between 1 (smallest fuzziness) and infinity. Its ideal value, however, is problem-dependent. Several distance measures d can be defined: the Euclidean distance d_E and the Mahalanobis distance d_M [11] are mostly used and are defined as:

$$d_E(\boldsymbol{x}_n, \boldsymbol{v}_c) = \|\boldsymbol{x}_n - \boldsymbol{v}_c\|^2 = \sqrt{\sum_{t=1}^{T} (v_{c,t} - x_{n,t})^2} \quad . \tag{5.4}$$

The algorithm starts from an initial set of membership values for the data set, expressed in matrix form as:

$$\boldsymbol{U}^0 = \left(1 - \frac{\sqrt{2}}{2}\right) \boldsymbol{U}_1 \; + \; \frac{\sqrt{2}}{2} \boldsymbol{U}_2 \quad , \tag{5.5}$$

with $\boldsymbol{U}_1$ a matrix with all elements equal to $1/C$ and $\boldsymbol{U}_2$ a matrix of randomly chosen elements. Next, the new cluster centres and memberships are computed using Eq.(5.2) and Eq.(5.3) . The procedure terminates when successive iterations do not further change significantly memberships and cluster centres, as calculated by Eq.(5.2) and Eq.(5.3). This procedure corresponds to the minimisation of the following objective function:

$${\sigma_W}^2 = \frac{1}{N} \sum_{c=1}^{C} \sum_{n=1}^{N} {u_{cn}}^m \, d(\boldsymbol{x}_n, \boldsymbol{v}_c) \quad , \tag{5.6}$$

which computes the within class variance over all clusters ${\sigma_W}^2$. In practice, a user-defined threshold for change in ${\sigma_W}^2$ determines when convergence is reached. The a priori determination of the fuzziness coefficient and the number of clusters are research topics often encountered in literature [18]. Although several heuristics are introduced, the result is often problem-dependent.

Preprocessing includes the transformation of each time series into its z-score as to avoid the clustering algorithm to classify the voxels based on signal amplitude, instead of signal shape. Finally PCA is performed to reduce data dimensionality.

5.2.2 Spatio-temporal clustering

To include spatial information to the clustering of time series, we introduce a spatial function h_{cn}, expressing the probability that a voxel $\boldsymbol{x}_n$ belongs to a cluster c, defined as:

$$h_{cn} = \sum_{k \in R\,(\boldsymbol{X}_n)} u_{ck} \quad . \tag{5.7}$$

Function h_{cn} is obtained by adding memberships in a predefined neighbourhood R (with size N_R) but only takes into account those neighbouring voxels exhibiting a strong membership to the cluster c of interest. For the calculation of the updated cluster centre in Eq.(5.2), the 'temporal' (i.e. based on time course data only) membership function u_{cn} is now replaced by a spatio-temporal membership ${u_{cn}}^{\mathrm{s}}$ as defined by:

$${u_{cn}}^{\mathrm{s}} = \frac{{u_{cn}}^{p}\ {h_{cn}}^{q}}{\sum_{k=1}^{C} {u_{kn}}^{p}\ {h_{kn}}^{q}} \quad . \tag{5.8}$$

Equation 5.8 describes a multiplicative model that expresses the membership in function of temporal (u_{cn}, Eq.(5.3)) and spatial (h_{cn}, Eq.(5.7)) components. Their relative influence is controlled by p and q, 2 exponents that vary independently. The denominator ensures each membership value for ${u_{cn}}^{\mathrm{s}}$ is in the range of 0 to 1. In this study simulated data sets were constructed for a broad CNR range (0.6 to 5) (see below) to assess the method's performance and compare to conventional FCM and to spatial smoothing followed by conventional FCM. The values of p and q for these analyses were determined by a preliminary analysis of several data sets. Results indicated that varying p and q independently did not significantly change compared to keeping p fixed (to unity) and varying q over a range (between 0.01 and 5), As the first solution yields much longer computation time, in the following we present results obtained using the latter. The neighbourhood's size N_{R} was restricted to three cases: $3 \times 3 \times 3$, $5 \times 5 \times 5$, and $7 \times 7 \times 7$. A $5 \times 5 \times 5$ neighbourhood for example covers an area of 2 voxels to the left and 2 to the right of the central voxel, in all three directions. No distance weighting was performed.

5.2.3 Data sets

Simulated data sets

A synthetic data set was constructed by adding realistic spatio-temporal activation patterns to background noise. Activation patterns were obtained from results of previous studies on visuospatial mental imagery [36, 37] (experiment is described in section 'Real data sets'). Typical fMRI auditory, imagery and visual activations with known intensity and spatial extent were extracted.
Fig. 5.1 shows a projection of the simulated activations on a flattened representation of the cortical sheet of the subject's brain (left and right hemisphere), mimicking auditory (green), imagery (blue), and visual (yellow) activation. Fig. 5.2 shows the corresponding time courses (a = auditory, b = imagery, c = visual). Time course period is 30 seconds (20 samples). These three different activations were selected in order to examine the method's performance under various circumstances and to take in consideration the variety of spatial structure normally

encountered in real fMRI data sets. The 'auditory' activation (transverse temporal gyrus and superior temporal sulcus/gyrus bilaterally) includes more contiguous regions as opposed to the 'imagery' activation (bilateral frontal and posterior parietal regions and occipito-temporal cortex regions), which includes more disperse and smaller regions.

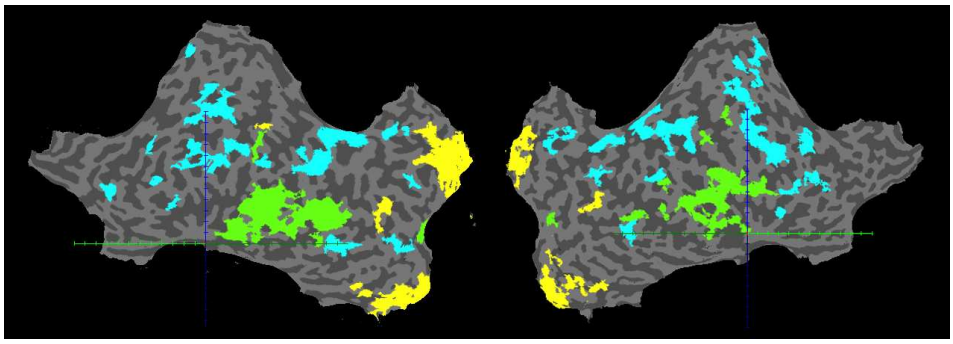

Fig. 5.1: *Projection of maps of inserted activation regions on inflated and flattened representation of the cortical sheet of the subject's brain (left and right hemisphere) for generated synthetic data sets: green = auditory; blue = imagery; yellow = visual.*

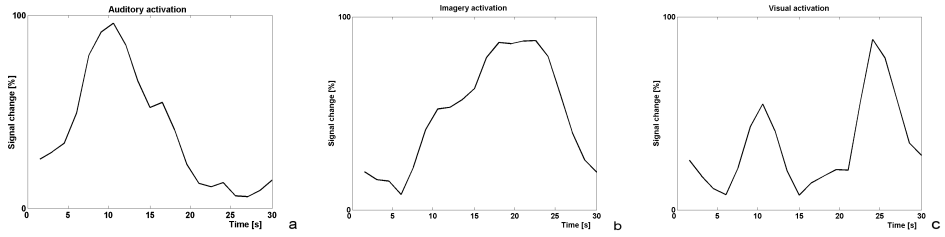

Fig. 5.2: *Time courses, corresponding to the activation regions of the synthetic data set of Fig. 5.1: (a) auditory, (b) imagery 1, and (c) visual. Time course period is 30 seconds.*

Realistic background noise originating from a null-data set, i.e. acquired from a subject who was resting during the entire imaging series, was added (Siemens Allegra, 3 Tesla, GE-EPI, TR = 1500 ms, TE = 46 ms, 32 slices, 64 × 64, voxel size = 3 × 3 × 3 mm^3). Relative amplitude of activation and background noise was controlled using the Contrast to Noise Ratio (CNR) which was varied in a range of 0.6 to 5. For the comparison with spatial smoothing and FCM, we spatially smoothed the synthetic data set using a Gaussian filter of FWHM (Full Width Half Maximum) of 4 mm and subsequently applied the conventional FCM algorithm. In this case, spatio-temporal FCM was applied using a weighting of spatial distance with a Gaussian function with equal FWHM.

Real data sets

Visuospatial mental imagery data are originating from functional measurements (see below), in which subjects were asked to create a mental representation of simple two-dimensional geometric figures based on a sequence of auditory instructions sequentially building up each figure. After a jittered delay, subjects had to mentally rotate the internally constructed image, in accordance with a visually presented target figure, rotated over an angle (40, 80, or 120 degrees), and had to indicate with a button press whether these were identical or mirror-inversed. This task involves a sequence of sensory, cognitive and motor processes which can take up to several seconds.

Three healthy, young subjects participated to the experiment. For each subject, several successive fMRI data runs were recorded in a single session: 3 for subject 'LM', 3 for subject 'AB', and 4 for subject 'CJ' (10 data sets in total). During each run, 18 trials were acquired each trial lasting 30 seconds. Functional scans consisted of 23 transversal slices acquired on a 3T Siemens Allegra (Siemens, Erlangen Germany) scanner using a single shot gradient-echo echo-planar imaging sequence (TE =30 ms, TR =1500 ms, matrix size = 64×64, voxel size = $3.5 \times 3.5 \times 4\,\mathrm{mm}^3$, 730 images). During the same session each subject underwent a high resolution T1 weighted anatomical scan (MDEFT sequence, voxel size =$1 \times 1 \times 1$ mm^3, 176 slices per slab, slice thickness = 1 mm, data matrix = 224 mm $\times$ 256 mm, TR = 7.92 s, TE = 2.4 ms).

The fMRI time series were subject to a series of pre-processing steps using BrainVoyagerQX (Brain Innovation, Maastricht, The Netherlands):

(1) Slice scan time correction was performed by resampling the time courses with sinc interpolation such that all voxels in a given volume represent the signal at the same point in time.

(2) Head movements were corrected automatically minimising the sum of squares of the voxel-wise intensity differences between each volume and the first volume of each run. Each volume was then resampled in three dimensional space according to the optimal parameters using trilinear interpolation.

(3) Temporal high pass filtering was performed to remove temporal drifts of a frequency below five cycles per run.

(4) After co-registration to the anatomical images collected in the same session functional volumes were projected into Talairach space [38].

(5) For each of the original 10 functional time series (3 subjects, 3 runs for subject 'AB' and 'LM' and 4 runs for subject 'CJ') an averaged data set was obtained considering an interval of 30 seconds around the onset of each auditory instruction (2 seconds pre stimulus onset and 28 seconds post stimulus) corresponding to 20 time points. This was done to increase the signal-to-noise

ratio of consistently task-related (CTR) processes, indispensable to have a reasonable FCM detection power of CTR processes [37].

5.2.4 Setting of parameters common to both FCM methods

Both conventional and spatio-temporal FCM require setting of specific parameters (fuzziness coefficient m, number of clusters C, threshold of convergence), which may have influence on the results of the analyses and the comparisons between methods. Their optimal setting was determined for synthetic and real data sets separately.

Simulated data sets

We constructed several synthetic data sets for CNR values in a range of 0.6 to 5 and analysed them subsequently.

- To determine an appropriate value of the fuzziness coefficient m, we applied both methods for a range of values of m between 1 and 3. After visual assessment of the topography and time course of the resulting clusters, 1.1 was chosen as an acceptable value for the fuzziness coefficient, which is in agreement with (data set dependent) reference values mentioned in literature [7, 18]. Higher values of m did not lead to detection of clusters out of the noise.

- Because only 3 types of activation were to be extracted out of the real noise data set, a number of clusters set to 8 proved to be sufficient. Indeed, initialising to higher or lower number of clusters yield to non optimal results. In case of lower number of clusters, some clusters were not always found (CNR dependent) and in case of higher number of clusters they were split into several clusters.

- We applied PCA to the data sets under investigation, capturing at least 90 % of its variance/covariance.

- In order to determine the influence of the initialization step, during which membership values and cluster centres are randomly chosen, we repeated the FCM decomposition multiple times and results were compared across multiple extractions. No significant difference in results was observed, indicating the robustness of the method to random initialization.

- Finally, we investigated the influence of spatio-temporal FCM's threshold of convergence on stability of the results, an important issue in view of the computation time. The threshold value appeared to have little influence on results for various CNR values if kept in a range of 10^{-2} to 10^{-5}. We thus used the less stringent value of 10^{-2}.

Real data sets

We determined the settings for real data sets by a preliminary analysis and inspection of one functional time series (subject 'LM', run 1).

- We selected 1.25 as an acceptable value for the fuzziness coefficient, which is in close agreement with reference values mentioned in literature [7, 18].
- The number of clusters was fixed to 13 for all the extractions.
- We applied PCA to the data sets under investigation and captured at least 90 % of its variance/covariance.
- A value of 10^{-2} as threshold of convergence led to stable results and an acceptable computation time for spatio-temporal FCM.
- Both methods also appeared robust to random initialization.

5.2.5 Selection and visualisation of clusters

Results of conventional and spatio-temporal FCM analysis of the data sets were assessed. We thereby focused on the clusters whose representative maps and time courses were clearly related to the imagery task, as described in a previous study [37]. For all data sets, maps obtained by both methods were superimposed to 3D anatomical images and/or projected on an inflated and flattened representation of the cortical sheet of the subject's brain. Overlay of cluster-maps corresponds to membership values in the range of 0.5 to 1. Hence, only those voxels clearly assigned to a single cluster are shown. The latter representation allows displaying in one picture the spatial topography of the clusters (as well as their corresponding time courses), thus providing a useful tool for their interpretation.

5.2.6 Assessment and comparison of results

We assessed quantitatively the results of conventional and spatio-temporal FCM by means of specific figures of merit.

For simulated data, we compared how well the results of both methods matched the simulated ground truth. For each method temporal assessment of a cluster's time course was obtained by the temporal correlation coefficient between the cluster centroid and the simulated time course.

Spatial assessment of the corresponding activation region was performed using Receiver Operating Characteristics (ROC) methodology. An ROC curve was constucted as a continuous plot of TPR (True Positive Rate) versus FPR (False Positive Rate) for a range of threshold values of membership uscn [39]. The latter acts as a rating parameter controlling the sensitivity of the detection procedure (i.e. FCM) [40]. To characterize detection accuracy we followed the approach of

Skudlarski et al. [41] and considered the mean of the ROC curve over a limited range of false-positive ratio between 0 and 0.1. By restricting the value of merit to low, but realistic false-positive rates (high thresholds only) we limited the scope of our analysis to the cases that are of primary interest in fMRI.

For real data, having unknown ground truth, the former approach, i.e. comparing how good results of both methods match the ground truth, is not feasible. In this study comparing the ability of methods to detect activation patterns corresponding to different processing stages of a visuo-spatial mental imagery experiment, we will visually compare corresponding cluster maps.

5.3 Results and discussion

5.3.1 Simulated data sets

Both conventional and spatio-temporal FCM decompositions allowed retrieval of all activations synthetically embedded in the noise background, for the CNR range under investigation. Results were consistent over all activations and runs.

5.3.1.1 Comparison to conventional FCM

Effect of relative influence of spatial information (parameter q)

Figure 5.3 shows the influence of q, the parameter expressing the relative influence of spatial information in the determination of the membership, on spatio-temporal FCM results in terms of the applied temporal and spatial figures of merit (CNR=1.2, $p = 1$, N_R=3 $\times$ 3 $\times$ 3). Results are differentiated for all types of simulated activations and are compared to conventional FCM.

For all activations, the spatial figure of merit is in favour of larger values of q, in which case spatio-temporal FCM always outperforms conventional FCM. The temporal figure of merit needs to be differentiated for the types of activation. For 'auditory' and 'visual' activation, an intermediate range of q exists within which spatio-temporal FCM outperforms conventional FCM. Optimal values of q lie between 0.5 and 2. The existence of the upper limit could be due to the fact that for excessive values of q the influence of border voxels on the membership may become too large. The latter yields a significant change in the cluster centre's time course. For 'imagery' activation however, we observe a descending trend of the temporal correlation as function of q, to such an extent that conventional FCM outperforms spatio-temporal FCM for larger values.

This anomalous result could be explained by the fact that the spatial pattern forming 'imagery' activation is more disperse compared to 'auditory' and 'visual' activation patterns, which are more contiguous and spatially clustered. Disperse clusters consist of many small regions and thus contain relatively large amounts

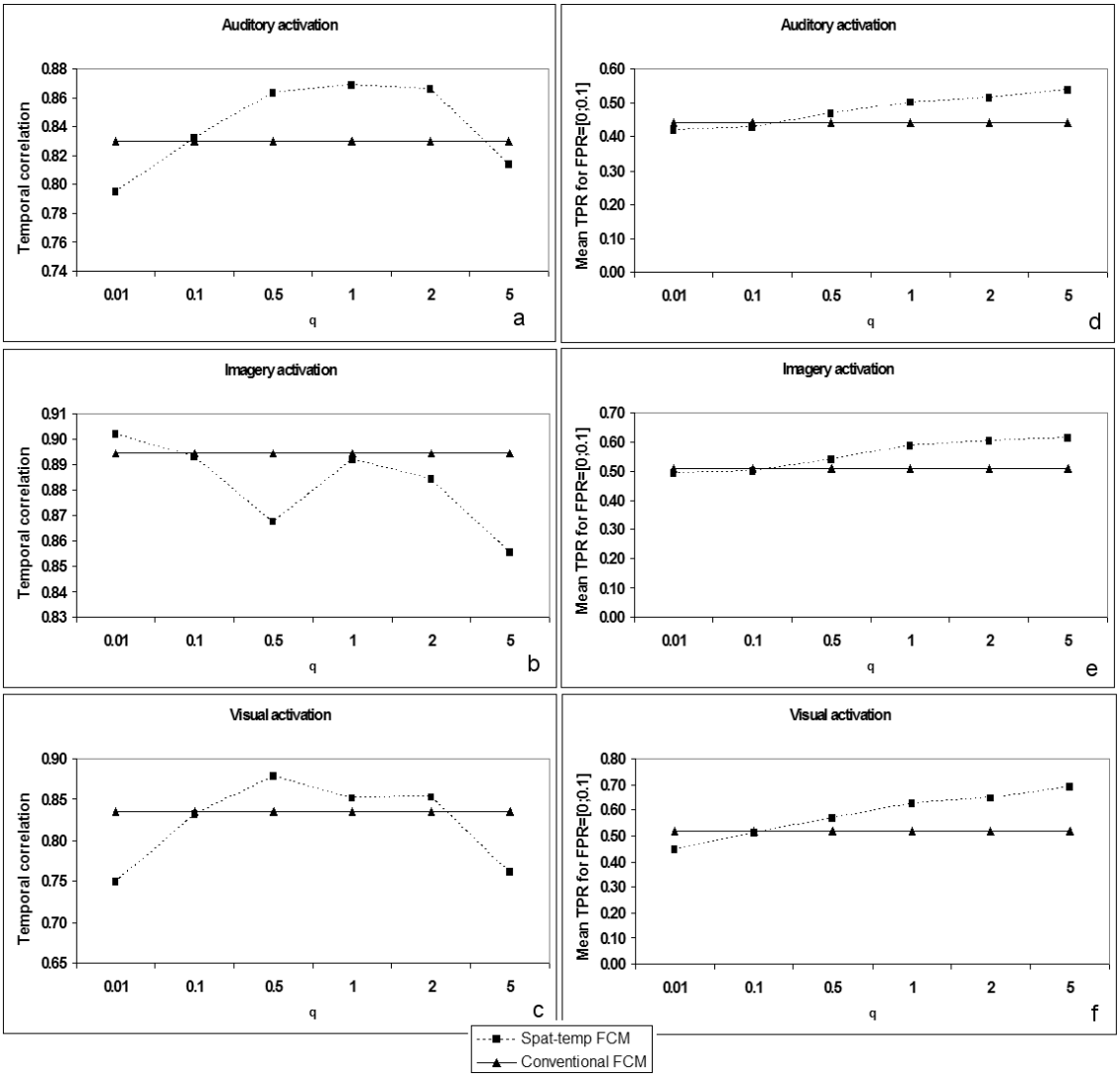

Fig. 5.3: *Influence of q (parameter expressing the relative influence of spatial information in the determination of the membership) on spatio-temporal FCM results of a synthetic data set (CNR=1.2, $p = 1$), compared to conventional FCM. Results are differentiated for the 3 types of activation. Figure of merit for temporal assessment is the temporal correlation coefficient (a, b, c) and figure of merit for spatial assessment is 'the mean TPR in a FPR range of 0 to 0.1' (d, e, f).*

of border voxels. The aforementioned effect thus arises even stronger, to such an extent that even for very low values of q the influence of border voxels on the membership becomes too large. Taking into account these findings, in this study we select the largest value of q which, for all types of activation, leads to the best assessment in terms of the temporal figure of merit. A value of $q = 1$ seems a good overall compromise.

Effect of neighbourhood size N_R

Figure 5.4 shows the influence of the neighbourhood size N_R on spatio-temporal FCM results in terms of the applied temporal (a) and spatial (b) figures of merit (CNR=1, $p = 1$, $q = 1$). Results are differentiated for all types of activation, and include the case 'no neighbourhood' ('NO NB'), i.e. conventional FCM.

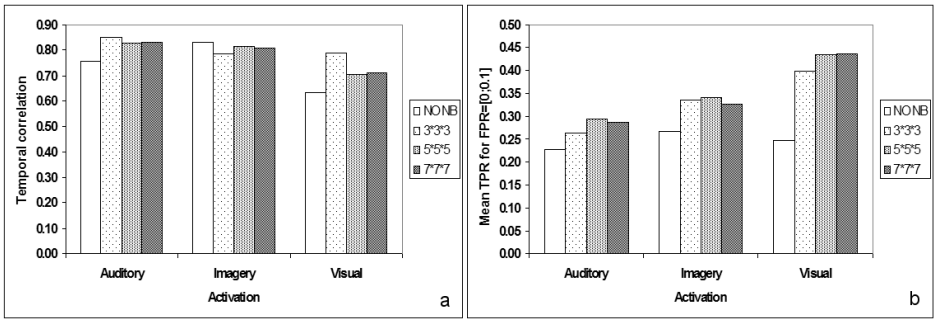

Fig. 5.4: *Influence of neighbourhood size on spatio-temporal FCM results of a synthetic data set (CNR=1, $p = 1$, $q = 1$), compared to conventional FCM (no neighbourhood, marked 'NO NB'). Figure of merit for temporal assessment is the temporal correlation coefficient (a) and figure of merit for spatial assessment is 'the mean TPR in a FPR range of 0 to 0.1' (b). Results are differentiated for the 3 types of activation.*

Both figures of merits show better performance for spatio-temporal FCM compared to conventional FCM, independent of the neighbourhood size, except for the temporal assessment of the 'imagery' activation. Results are not unambiguous however concerning the influence of the neighbourhood size on the figures of merit. The temporal figure of merit only suggests improvement in detection accuracy for 'auditory' and 'visual' activation, not for 'imagery' activation. This could again be explained by the dispersiveness of the 'imagery' region.

However, no general inference about the optimal neighbourhood size can be made, due to the limited number of activation types involved in this study and due to the fact that only one realistic noise data set was used to construct the different simulated data sets (each for a different CNR). In view of the smaller computation time a $3 \times 3 \times 3$ neighbourhood is selected in this study.

Effect of CNR

Figure 5.5 shows the influence of CNR on spatio-temporal FCM results in terms of the temporal (a) and spatial (b) figures of merit ($p = 1$, $q = 1$, $N_{\mathrm{R}} = 3 \times 3 \times 3$). Results are shown for 'auditory' activation (analogous results were obtained for the 'imagery' and 'visual' activations) and are compared to conventional FCM.

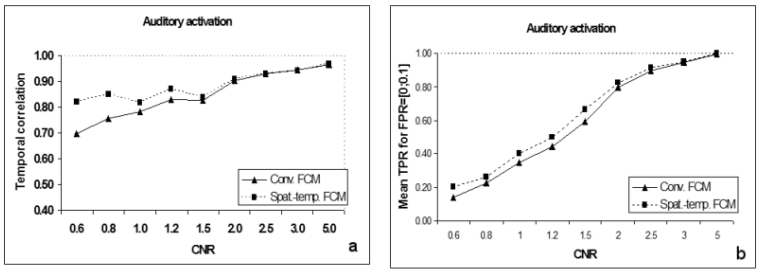

Fig. 5.5: *Influence of CNR on spatio-temporal FCM results of a synthetic data set ($p = 1$, $q = 1$), compared to conventional FCM. Figure of merit for temporal assessment is the temporal correlation coefficient (a) and figure of merit for spatial assessment is 'the mean TPR in a FPR range of 0 to 0.1' (b). Results are shown for auditory activation (analogous for imagery and visual activation).*

Both figures of merit show that fuzzy clustering techniques perform better for data characterised by higher CNR values, as already illustrated by Fadili et al. [40]. For data sets characterised by lower CNR values (i.e. up to 2.5), both figures of merit indicate that spatio-temporal FCM outperforms conventional FCM, a finding in line with the conclusions of Chuang et al. [32]. For higher CNR values, the assessment of both methods in terms of the figures of merit becomes comparable.

5.3.1.2 Difference with spatial smoothing

Tab. 5.1 compares results between the two approaches in terms of the discussed figures of merit and is differentiated for all activations (data set with CNR=1). A first approach (marked '1') applies our spatio-temporal FCM algorithm with a Gaussian neighbourhood window of FWHM of 4 mm. A second approach (marked '2') consists in spatially smoothing the synthetic data set using a Gaussian filter of equal FWHM and subsequently applying the conventional FCM algorithm. Table 1 also shows results in case stronger smoothing is applied for the latter approach, i.e. using a Gaussian filter with FWHM of 8 mm (marked '')).

Activation type	Auditory			Imagery			Visual		
Method	1	2	3	1	2	3	1	2	3
mean TPR in FPR range [0;0.1]	0.38	0.38	0.39	0.48	0.31	0.19	0.46	0.42	0.55
temporal correlation coefficient	0.84	0.80	0.75	0.87	0.70	0.59	0.80	0.82	0.81

Tab. 5.1: *Table 1. Comparison between two FCM-based methods dealing with noise for the analysis of a synthetic data set (CNR=1). The comparison is made in terms of 2 figures of merits: 'the mean TPR in a FPR range of 0 to 0.1' (spatial assessment) and the temporal correlation coefficient (temporal assessment). Results are differentiated for all types of activation. Spatio-temporal FCM (NB=3 x 3 x 3, $p = 1$, $q = 1$, Gaussian neighbourhood with FWHM=4 mm) is marked '1'. Spatial smoothing with a Gaussian filter, followed by conventional FCM is marked '2' in case a FWHM of 4 mm is used. Spatial smoothing with a Gaussian filter, followed by conventional FCM is marked '3' in case a FWHM of 8 mm is used.*

The comparison indicates that spatio-temporal FCM clearly outperforms the preliminary smoothing followed by conventional FCM in the case of the 'imagery' activation, and, to a smaller extent, in the other cases. Smoothing more severely the data does not result in a clear improvement for 'auditory' and 'visual' activation, and even results in deterioration for 'imagery' activation. Spatial smoothing as a technique to reduce noise in fMRI data sets suffers from major shortcomings, like blurring and loss of low-amplitude signals. These shortcomings manifest themselves particularly in case of small, disperse activation regions, like the 'imagery' activation. The ensuing worsening of detection accuracy, especially in case of low CNR data sets, probably explains the outperformance (in terms of both figures of merit) of our method, incorporating spatial information into the procedure. This is line with findings of FCM-based approaches adding spatial information directly to the time courses [34] and findings of model-driven approaches, combining temporal and spatial information [29, 30].

5.3.2 Real data sets

5.3.2.1 Comparison to conventional FCM

Both conventional and spatio-temporal FCM decompositions allowed highlighting the contribution of distinct networks of areas to the sequential stages of the task. Results were consistent over activations and runs.

Figure 5.6 shows the time courses of cluster centroids found by conventional FCM (solid line) and spatio-temporal FCM (dash dot line). The amplitudes are normalized. The baseline level is determined by the vertical line indicating the onset of the different stages of the experiment. Figure 5.7 shows a projection of thresholded and binarised membership maps of the corresponding activation regions on an in-

flated and flattened representation of the cortical sheet of the subject's brain, for a typical visuospatial mental imagery data set (subject 'AB', run 4, left and right hemisphere). The different stages of the experiment are indicated. Regions resulting from conventional FCM analysis are presented in yellow, regions resulting from spatio-temporal FCM in red. Regions found by both methods are shown in orange.

Several patterns of activation are found: (a) reflects the activation of the auditory cortex at the beginning of the trial. Subsequently, response is found representative of brain activation during the performance of a mental imagery task. As suggested in a previous study [37] the subdivision in (b) and (c) indicates a sequential involvement of corresponding cortical networks in early and late stages of this imagery task. Response related to visual presentation of the target stimulus is represented in (d), while (e) is dominated by subsequent motor response. Motor-related activity corresponds to a right hand button press.

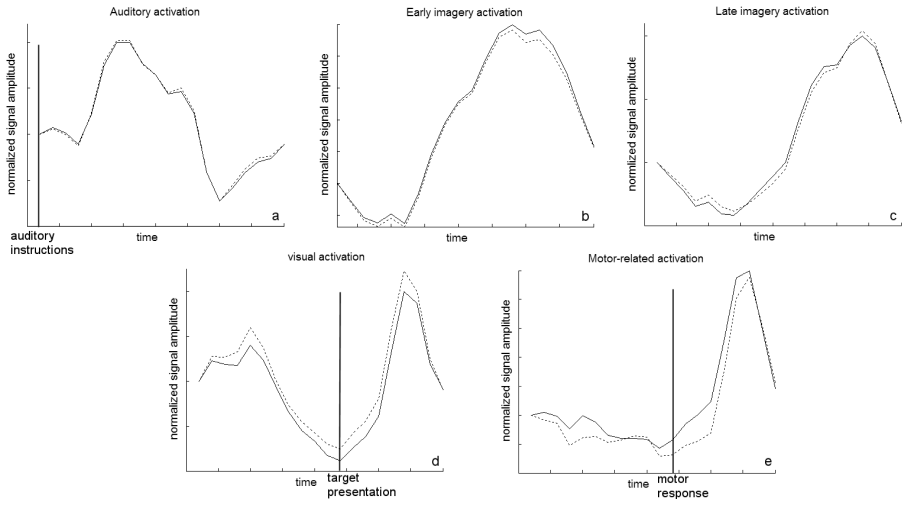

Fig. 5.6: *Time courses of cluster centres found by conventional FCM (solid line) and spatio-temporal FCM (dash-dot line): (a) auditory, (b) early imagery, (c) late imagery, (d) visual, (e) motor. Time course period is 30 seconds. The baseline level is determined by the vertical line indicating the onset of the different stages of the experiment.*

Both methods perform clustering on time course features, yielding a large temporal agreement across methods, shown in Fig. 5.6. Fig. 5.7 indicates that both methods allowed highlighting the contribution of distinct networks of areas to the sequential stages of the task, with good spatial agreement across methods. A closer visual comparison between spatial results obtained by both methods suggests that an even finer determination of physiologically interesting clusters is achieved by spatio-temporal FCM. Spatio-temporal FCM is more successful (compared to con-

ventional FCM) in the detection of following regions:

(1) Spatio-temporal FCM includes the fringes to regions detected by conventional FCM.

(2) Spatio-temporal FCM reveals small detached regions belonging to a detected cluster, but left unrecognised by conventional FCM.

(3) Spatio-temporal FCM favours more contiguous regions, as compared to conventional FCM. Nearby regions, left distinct by conventional FCM are often interconnected by spatio-temporal FCM by adding intermediate regions, thus resulting in a more contiguous area.

The regions additionally found by spatio-temporal FCM are characterised by lower membership values compared to the spatio-temporal FCM membership values of regions found by both methods. It may be assumed that the degree to which a brain area is activated for the performance of a processing stage of a task is expressed by the FCM membership of a voxel to the corresponding cluster. Moreover, it is a reasonable assumption that the standard deviation of the noise is not changing significantly within each of the detected activated regions. This infers that the regions, additionally detected by spatio-temporal FCM are characterised by lower CNR values compared to the regions found by both methods, a finding which is in agreement with the results of the simulated data sets.

5.4 Conclusions

Spatio-temporal FCM is an approach for the analysis of fMRI data sets that combines temporal and spatial information for the assignment of a voxel to a specific cluster. We have demonstrated the validity of this approach by analysing realistic simulated data sets and by comparing results to conventional FCM. This comparison showed that adding spatial information to the fuzzy clustering algorithm improves its detection accuracy. In practical application, the advantage of using spatio-temporal rather than conventional FCM may depend on the nature of the specific data set and on the settings of various parameters. Therefore, we examined the influence of the most relevant of these parameters and evaluated the differences between the two approaches for different topographies of activations and for a range of functional CNR values, which are normally encountered in fMRI analysis. Spatio-temporal FCM performed significantly better than conventional FCM for the analysis of data sets characterised by low CNR data. This property resulted in the detection of region's fringes and weaker activated small regions. Moreover it led to the detection of more contiguous regions. However, in order to confirm the generality of these findings, more research is necessary.

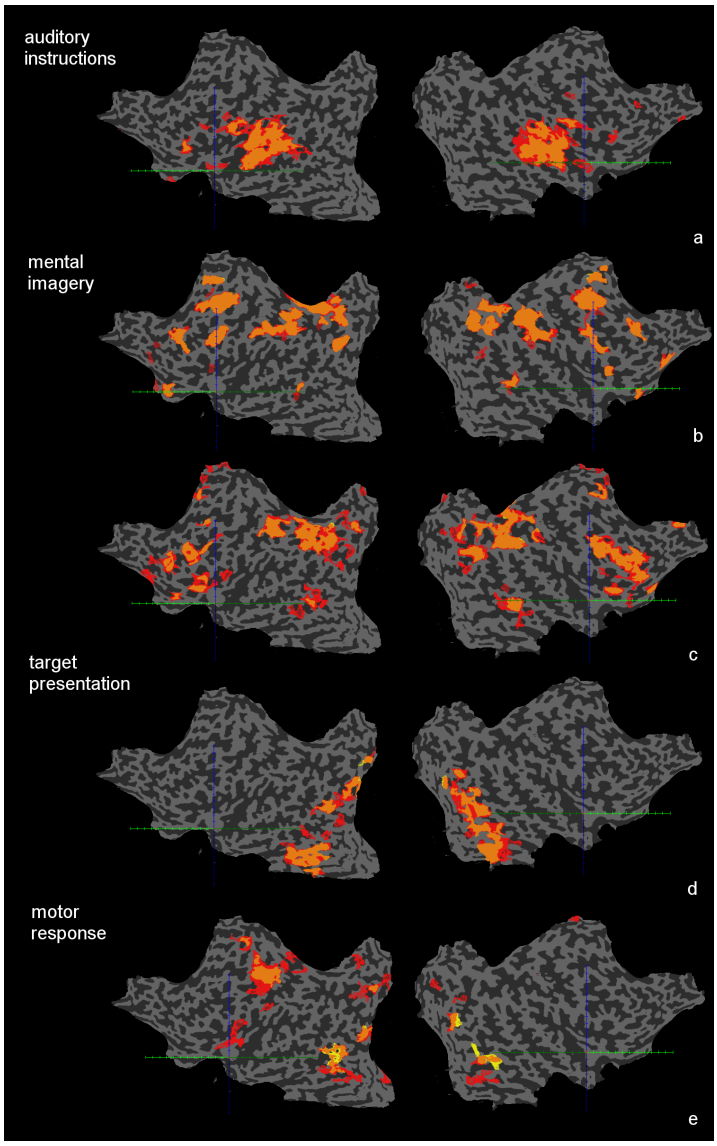

Fig. 5.7: *Projection of binarised thresholded membership maps of activation regions on inflated and flattened representation of the cortical sheet of the subject's brain for a typical data set (subject 'AB', run 4, left and right hemisphere) for the applied visuospatial mental imagery task. The data set was analysed with conventional FCM (yellow) and spatio-temporal FCM (red), regions found by both methods are shown in orange. Activation regions found are (a) auditory, (b) early imagery, (c) late imagery, (d) visual, (e) motor. Motor-related activity corresponds to a right hand button press. The different stages of the experiment are indicated.*

Bibliography

Bibliography

[1] K J Friston et al. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4), 1994.

[2] JR Duann et al. Single-trial variability in event-related BOLD signals. *Neuroimage*, 15(4), 2002.

[3] G K Aguirre et al. The variability of human, BOLD hemodynamic responses. *Neuroimage*, 8(4):360–369, Nov 1998.

[4] M J McKeown et al. Analysis of fMRI data by blind separation into independent spatial components. *Hum Brain Mapp*, 6(3):160–188, 1998.

[5] C Goutte et al. On clustering fMRI time series. *Neuroimage*, 9(3):298–310, Mar 1999.

[6] L A Zadeh. *Fuzzy Sets and their Application to Pattern Classification and Clustering Analysis*. Academic Press, New-York, 1977.

[7] G Scarth et al. Detection of novelty in functional images using fuzzy clustering. In *Proceedings of the 3rd Meeting ISMRM*, pages 238–..., Nice, France, 1995.

[8] R Baumgartner et al. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part I: reproducibility. *J Magn Reson Imaging*, 7 (6):1094–1101, Nov 1997.

[9] E Moser et al. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part II: quantification. *J Magn Reson Imaging*, 7(6):1102–1108, Nov 1997.

[10] X Golay et al. A new correlation-based fuzzy logic clustering algorithm for fMRI. *Magn Reson Med*, 40(2):249–260, Aug 1998.

[11] P C Mahalanobis. On generalized distance in statistics. In *Proceedings National Institute of Sciences India*, volume 12, pages 49–55, 1936.

[12] R Baumgartner et al. Quantification in functional magnetic resonance imaging: fuzzy clustering vs. correlation analysis. *Magn Reson Imaging*, 16(2): 115–125, 1998.

[13] R Baumgartner et al. Comparison of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis. *Magn Reson Imaging*, 18(1):89–94, Jan 2000a.

[14] LO Hall et al. A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. volume 3, pages 672–682, IEEE Transactions on neural networks, 1992.

[15] Y Lu et al. A split-merge-region based region-growing method for fMRI activation detection. *Human Brain Mapping*, 22(4):271–279, 2004.

[16] A Meyer-Baese et al. Comparison of two exploratory data analysis methods for fMRI: unsupervised clustering versus independent component analysis. *IEEE Trans Inf Technol Biomed*, 8(3):387–398, Sep 2004.

[17] E Dimitriadou et al. A quantitative comparison of functional MRI cluster analysis. *Artif Intell Med*, 31(1):57–71, May 2004.

[18] M J Fadili et al. On the number of clusters and the fuzziness index for unsupervised FCA application to BOLD fMRI time series. *Med Image Anal*, 5(1):55–67, 2001.

[19] H Jahanian et al. Roc-based determination of number of clusters for fmri activation detection. In *Proceedings of SPIE*, volume 5370, 2004.

[20] A Baune et al. Dynamical cluster analysis of cortical fMRI activation. *Neuroimage*, 9(5):477–489, May 1999.

[21] R Baumgartner et al. Resampling as a cluster validation technique in fMRI. *J Magn Reson Imaging*, 11(2):228–231, Feb 2000b.

[22] U Moller et al. How to avoid spurious cluster validation? A methodological investigation on simulated and fMRI data. *Neuroimage*, 17(1):431–446, Sep 2002.

[23] W F Auffermann et al. Cluster significance testing using the bootstrap. *Neuroimage*, 17(2):583–591, Oct 2002.

[24] M Jarmasz and R L Somorjai. Exploring regions of interest with cluster analysis (EROICA) using a spectral peak statistic for selecting and testing the significance of fMRI activation time-series. *Artif Intell Med*, 25(1):45–67, May 2002.

[25] C Windischberger et al. Fuzzy cluster analysis of high-field functional MRI data. *Artif Intell Med*, 29(3):203–223, Nov 2003.

[26] C Goutte et al. Feature-space clustering for fMRI meta-analysis. *Hum Brain Mapp*, 13(3):165–183, Jul 2001.

[27] O Simon et al. Automatized clustering and functional geometry of human parietofrontal networks for language, space, and number. *Neuroimage*, 23(3): 1192–1202, Nov 2004.

[28] H Jahanian et al. Functional magnetic resonance imaging activation detection: fuzzy cluster analysis in wavelet and multiwavelet domains. *J Magn Reson Imaging*, 22(3):381–389, Sep 2005.

[29] X Descombes et al. Spatio-temporal fMRI analysis using Markov random fields. *IEEE Trans Med Imaging*, 17(6):1028–1039, Dec 1998.

[30] S J Kiebel et al. Anatomically informed basis funtcions . *Neuroimage*, 11: 656–667, 2000.

[31] M N Ahmed et al. A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. *IEEE Trans Med Imaging*, 21(3): 193–199, Mar 2002.

[32] K S Chuang et al. Fuzzy c-means clustering with spatial information for image segmentation. *Comput Med Imaging Graph*, 30(1):9–15, Jan 2006.

[33] S Shen et al. MRI fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization. *IEEE Trans Inf Technol Biomed*, 9(3):459–467, Sep 2005.

[34] R L Somorjai et al. A novel, direct spatio-temporal approach for analyzing fMRI experiments. *Artif Intell Med*, 25(1):5–17, May 2002.

[35] J C Bezdek. FCM: the fuzzy C-means algorithm. *Computers and Geosciences*, 10:191–203, 1984.

[36] E Formisano et al. Tracking the mind's image in the brain I: time-resolved fMRI during visuospatial mental imagery. *Neuron*, 35(1):185–194, Jul 2002.

[37] A Smolders et al. Dissecting cognitive stages with time-resolved fMRI data: a comparison of fuzzy clustering and independent component analysis. *Magn Reson Imaging*, May 2007.

[38] J Talairach and P Tournoux. *Co-planar Stereotaxic Atlas of the Human Brain.* New York: Thieme Medical Publications, 1988.

[39] R T Constable et al. An ROC approach for evaluating functional brain MR imaging and postprocessing protocols . *Magnetic Resonance in Medicine*, 34 (1):57–64, 1995.

[40] M J Fadili et al. A multistep unsupervised fuzzy clustering analysis of fMRI time series. *Human Brain Mapping*, 10:160–178, 2000.

[41] P Skudlarski et al. ROC analysis of statistical methods used in functional MRI: individual subjects . *Neuroimage*, 9(3):311–329, 1999.

Chapter 6

General conclusions

This dissertation deals with data-driven methods for the analysis of BOLD fMRI data sets, resulting from a complex, cognitive task.

In the first part of this dissertation, we elaborately reviewed the underlying physiological principles, the experimental design, and the analysis techniques, as to illustrate the shortcomings of actual approaches for the analysis of complex, cognitive data sets. We elucidated the BOLD fMRI technique to map brain functionality. More specifically, we elaborated on the relationship between the measured haemodynamic effect and its underlying neuronal activity. We showed that the complexity of this relationship, as well as many confounding factors severely complicate the interpretation of the measured data. We emphasized that the emergence of complex cognitive tasks, such as mental imagery, especially obliged an appropriate choice of experimental design and subsequent analysis technique. We illustrated that an event-related, time-resolved data acquisition technique, combined with a data-driven analysis method yields the best approach to retrieve the hierarchical processing stages of such tasks. We focused on two successful techniques: the fuzzy clustering method (FCM) and spatial independent component analysis (sICA). We emphasised that, for the analysis of a complex data set, an appropriate selection of method-specific parameters and the correct validation of the results is difficult to accomplish.

In the second part of this dissertation, we introduced new approaches and techniques based on FCM and sICA, as to overcome the aforementioned shortcomings. We determined guidelines for the selection between methods as well as for their

use when analysing complex, cognitive data sets. We also incorporated neighbourhood information in the FCM algorithm. We applied techniques for the assessment and comparison of these approaches and algorithms and provided time-resolved, event-related mental imagery data sets. Therefore, we introduced a newly devised visuospatial mental imagery experiment, in which subjects were asked to online mentally construct abstract geometric figures by juxtaposing six sequentially auditory presented stimuli. After the construction phase, a target stimulus was visually presented and evaluated.

We evaluated the proposed approaches and algorithms in two studies. In a first study, we determined the influence of method-specific parameter setting of FCM and sICA on detection accuracy. We illustrated that both techniques allow to highlight the contribution of distinct networks of brain regions to the sequential stages of the task (auditory perception, mental imagery, and behavioural response). We demonstrated that, on averaged time series, FCM leads to a richer decomposition of activation patterns. In particular, FCM found additional clusters presumably related to the early and late processing stages of all aspects of the tasks (e.g. early and late imagery). Furthermore, we performed a quantitative comparison between the methods. The high within- and between subject consistency of clusters/ICs, together with the anatomical layout and temporal profile suggested a neurophysiologically meaningful separation. Spatial and temporal correlations of cluster/IC maps and time courses showed a good agreement between results of both methods, suggesting that the partitions of the time series obtained using both methods are comparable. Finally, the proposed FCM approach was incorporated in the Brainvoyager QX analysis software.

In a second study, we incorporated neighbourhood information into the conventional FCM algorithm. We constructed realistic simulated data sets and analysed them using conventional and spatio-temporal FCM. We demonstrated the validity of the spatio-temporal FCM approach by a quantitative comparison between results obtained by both approaches. We illustrated that, in practical applications, the advantage of using spatio-temporal FCM rather than conventional FCM depends on the nature of the specific data set, and the setting of various parameters. We showed that, for different topographies of activations and for typical fMRI data sets, improvements were most significant for data sets characterised by low CNR values.

Although the proposed FCM and sICA based approaches and techniques were able to dissect the main sequential processing stages of a cognitive task, a finer discrimination of neuro-cognitive stages may still be achieved. In the analysis of the visuospatial mental imagery data set, for example, we were only able to distinguish between an early and a late imagery stage. We were not able to link these stages to a particular sub-task of the imagery experiment. Future investigation aims at highlighting networks corresponding to all imagery sub-processes, such as building and keeping in mind of the figure, mental rotation of the constructed

figure, and subsequent mental comparison to the presented target. A successful approach however does not merely rely on the experimental design and analysis techniques, but also on the acquisition techniques.

Nederlandse samenvatting

Dit proefschrift handelt over data-gedreven methodes voor de analyse van complexe datasets. De gemeten signalen zijn bekomen met functionele Magnetische Resonantie Beeldvorming (fMRI) en gebaseerd op het in het bloed gemeten zuurstofniveau (BOLD). Het onderzoekt en identificeert de spatiale lay-out en de sequentie van hersenactivaties bij de uitvoering van een cognitieve taak. De relatie tussen het gemeten hemodynamisch effect en het onderliggend neuronaal effect is echter complex en bovendien slecht begrepen. Bovendien is het gemeten signaalverschil zwak en onderhevig aan niet-lineariteiten. Tenslotte wordt de interpretatie van de resultaten bemoeilijkt door interfererende fenomenen en de aanwezigheid van verschillende ruisbronnen. Deze beperkingen worden bijzonder doorslaggevend bij de analyse van complexe, cognitieve datasets. Er bestaan verschillende manieren om met deze beperkingen om te gaan, zoals een aangepaste keuze van experimenteel ontwerp, acquisitie techniek en analyse methode. In deze dissertatie ligt de nadruk op het laatste aspect, de analyse methode.

Initieel werden hypothese-gedreven analysetechnieken toegepast. Deze specifiëren a priori een spatiaal-invariant model van de hemodynamische respons (HR). De mate waarin dat model geldt wordt voor elk voxel bepaald met behulp van statistische methodes. De veronderstelling dat de HR spatiaal invariant is leidt tot sub-optimale resultaten wanneer complexe datasets worden geanalyseerd. Deze datasets brengen immers de activatie van uitgebreide netwerken binnen de hersenen met zich mee. De betrokken HRs verschillen sterk van elkaar. Bovendien is de HR binnen elk gebied onderworpen aan een grote variabiliteit bij de uitvoering van verschillende metingen. Hypothese-gedreven methodes tenslotte maken geen gebruik van interacties tussen voxels. Om al deze redenen werd een alternatieve, data-gedreven aanpak naar voor geschoven. Dit leidde tot de toepassing van de Vage Clustering Methode (FCM) en de spatiale Onafhankelijke Componenten

Analyse (sICA). Beide methodes ontleden de data in een set van spatio-temporele modes, zonder a priori sterke voorwaarden op te leggen aan de signaalvorm. Beide methodes hebben echter een complementaire kijk op de data. Spatiale ICA bekijkt de data vanuit een spatiaal perspectief, d.w.z. dat de data behandeld worden als een spatiale verdeling van tijdsseries, terwijl FCM de data bekijkt vanuit een temporeel perspectief, d.w.z. dat de data behandeld worden als een sequentie van volumes. FCM verdeelt de tijdsseries in enkele types van activatie, die de onderliggende neuronale activiteit weergeven. Elke groep of cluster wordt gekarakteriseerd door een spatiaal gebied en een bijhorende tijdsserie, het cluster centrum. FCM drukt de overeenkomst tussen de tijdsserie van een voxel en een bepaald cluster centrum uit aan de hand van een zogenaamd lidmaatschap van die cluster. Spatiale ICA is een informatie-theoretische benadering die statistische onafhankelijkheid tussen de componenten nastreeft. De methode modelleert de data als een lineaire combinatie van onafhankelijke componenten (ICs), die de onderliggende neuronale activiteit weergeven.
Beide methodes vertonen specifieke tekortkomingen wanneer complexe, cognitieve datasets worden geanalyseerd. Het is voor deze datasets immers bijzonder moeilijk om de methode-specifieke parameters weloverwogen te bepalen alsook om de resultaten correct te valideren. Bovendien kunnen beperkingen in de inherente nauwkeurigheid van de detectie van de methodes een correcte dissectie van de verschillende verwerkingsstadia van complexe, cognitieve taken in de weg staan.

In deze dissertatie worden twee manieren voorgesteld, gebaseerd op FCM en sICA, om deze tekortkomingen te ondervangen. Vooreerst worden er richtlijnen opgesteld betreffende de keuze tussen de beide methodes en de instelling van hun parameters in de context van een complexe, cognitieve taak. Vervolgens stellen we een krachtiger FCM algoritme voor dat spatiale informatie mee in beschouwing neemt. Deze benadering houdt rekening met de verwachting dat taak-geïnduceerde neuronale activaties en BOLD responsies gelijkaardige signaalveranderingen produceren in spatiaal aanéén gesloten gebieden, die zich uitstrekken over verschillende millimeters.

Deze benaderingen en algoritmes worden geëvalueerd en vergeleken op basis van de nauwkeurigheid en consistensie van de detectie, de robuustheid, het gemak waarmee de parameters ingesteld kunnen worden en de valideerbaarheid en interpreteerbaarheid van de resultaten. Om deze criteria te kunnen beoordelen worden geschikte datasets voorgesteld. In verband daarmee wordt een nieuw ontworpen experiment m.b.t. visuospatiale mentale inbeelding voorgesteld. Dit experiment resulteert in tijds-geresolveerde mentale chronometrie fMRI datasets.

Er bestaan twee manieren om de beoordeling van deze criteria te bewerkstelligen. Om deze criteria *kwantitatief* te evalueren zijn simulaties en testen op *synthetische* datasets onontbeerlijk. De evaluatie aan de hand van *reële* datasets daarentegen kan alleen *kwalitatief* uitgevoerd worden. Bevindingen van eerder verricht onderzoek, met andere analysetechnieken en op dezelfde of soortgelijke datasets, biedt

echter een referentiekader aan. Beide manieren worden toegepast in deze dissertatie, in twee verschillende studies.

Een eerste studie stelt empirische richtlijnen op voor de keuze tussen beide methodes, FCM en sICA, en voor hun gebruik in de praktijk. Beide algoritmes worden geëvalueerd en vergeleken aan de hand van reële datasets, bekomen in de context van het experiment m.b.t. visuospatiale mentale inbeelding. Een kwalitatieve beoordeling van de valideerbaarheid en interpreteerbaarheid van fysiologisch betekenisvolle componenten, alsook de robuustheid van de methode wordt uitgevoerd. Daarnaast worden de nauwkeurigheid van detectie en consistentie van de gevonden tijdsseries en spatiale gebieden binnen eenzelfde proefpersoon en tussen verschillende proefpersonen kwantitatief geëvalueerd en vergeleken. Daartoe wordt respectievelijk de temporele en spatiale correlatie bepaald. In deze studie wordt eveneens nagegaan in welke mate de resultaten van deze beoordelingscriteria beïnvloed worden door een aantal factoren. Enerzijds worden de methode-specifieke parameterinstellingen beschouwd, anderzijds de invloed van het uitmiddelen van de - tijdens opeenvolgende proeven gemeten - signalen vooraleer ze ter analyse aan te bieden.

Een tweede studie introduceert het 'spatio-temporele' FCM algoritme dat informatie van de naburige meetpunten in het detectieproces opneemt. Het verheldert de condities en de parameter-instellingen waaronder spatio-temporele FCM 'conventionele' FCM overtreft op het gebied van de nauwkeurigheid van detectie. Daarom worden realistische gesimuleerde datasets aangemaakt met verschillende types van activatie, die synthetisch ingebracht zijn in een realistische ruisachtergrond. Deze datasets worden gegenereerd voor het in fMRI-analyse gebruikelijke bereik van de 'Contrast-Ruis-Verhouding' (CNR). Er wordt een kwantitatieve vergelijking uitgevoerd van de activatiegebieden en bijhorende tijdsseries, bekomen via beide methodes. Dit wordt respectievelijk gerealiseerd aan de hand van een techniek bekend als 'ontvanger bedrijfskarakteristieken' (ROC) en de temporele correlatiecoëfficiënt.
Tenslotte worden deze bevindingen, bekomen uit gesimuleerde datasets, toegepast op de analyse van de datasets bekomen uit het experiment m.b.t. visuospatiale mentale inbeelding. Deze datasets worden geanalyseerd met beide methodes om de verschillen tussen de overeenkomstige topologische structuren te bepalen en te interpreteren.