Functional magnetic resonance imaging: an intermediary between behavior and neural activity

A Thesis Submitted to the
College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the degree of Doctor of Philosophy
in the Division of Biomedical Engineering
University of Saskatchewan
Saskatoon

By
Vasily A. Vakorin

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Division of Biomedical Engineering

Engineering Building

57 Campus Drive

University of Saskatchewan

Saskatoon, Saskatchewan

Canada

S7N 5A9

Abstract

Blood oxygen level dependent (BOLD) functional magnetic resonance imaging is a non-invasive technique used to trace changes in neural dynamics in reaction to mental activity caused by perceptual, motor or cognitive tasks. The BOLD response is a complex signal, a consequence of a series of physiological events regulated by increased neural activity. A method to infer from the BOLD signal onto underlying neuronal activity (hemodynamic inverse problem) is proposed in Chapter 2 under the assumption of a previously proposed mathematical model on the transduction of neural activity to the BOLD signal. Also, in this chapter we clarify the meaning of the neural activity function used as the input for an intrinsic dynamic system which can be viewed as an advanced substitute for the impulse response function. Chapter 3 describes an approach for recovering neural timing information (mental chronometry) in an object interaction decision task via solving the hemodynamic inverse problem. In contrast to the hemodynamic level, at the neural level, we were able to determine statistically significant latencies in activation between functional units in the model used. In Chapter 4, two approaches for regularization parameter tuning in a regularized-regression analysis are compared in an attempt to find the optimal amount of smoothing to be imposed on fMRI data in determining an empirical hemodynamic response function. We found that the noise autocorrelation structure can be improved by tuning the regularization parameter but the whitening-based cri-

terion provides too much smoothing when compared to cross-validation. Chapter 5 illustrates that the smoothing techniques proposed in Chapter 4 can be useful in the issue of correlating behavioral and hemodynamic characteristics. Specifically, Chapter 5, based on the smoothing techniques from Chapter 4, seeks to correlate several parameters characterizing the hemodynamic response in Broca's area to behavioral measures in a naming task. In particular, a condition for independence between two routes of converting print to speech in a dual route cognitive model was verified in terms of hemodynamic parameters.

Acknowledgements

I am grateful for collaboration and discussions with Dr. Gordon Sarty and Dr. Ron Borowsky. Also, my thanks go to Jeff Bird for using my software to extract temporal and spatial parameters of the hemodynamic response used in the analysis of Chapter 5. In addition, I would like to thank Jacqueline Cummine for neuroanatomy help, in particular, for masking regions of interests and correlating behavioral and hemodynamic parameters in Chapter 5. I would also like to acknowledge that all of fMRI data used in the thesis came from cognitive science experiments designed and conducted by Dr. Ron Borowsky, Dr. Gordon Sarty and Jacqueline Cummine (Chapters 3,5), Dr. Ron Borowsky and Dr. Gordon Sarty(Chapter 4) and by Dr. Ron Borowsky, Dr. Gordon Binsted, Dr. Gordon Sarty and Sofia Vorontsova (Chapter 2). I would like to extend my appreciation to Dr. Inna Molodtsova for continuous support during my staying in Saskatoon. Also, my special thanks go to Mikhail Vasilyevich Lomonosov for inspiration. Finally, I am grateful to Dr. Olga Krakovska for being in my life and to my parents for their love.

To my parents

Contents

List of Tables

List of Figures

List of Abbreviations

BOLD	bold oxygen level dependent
CBF	cerebral blood flow
CBV	cerebral blood volume
DCM	dynamic causal model
EEG	Electroencephalography
EET	epoxyeicosatrienoic acids
EM	expectation-maximization
FDA	functional data analysis
FID	free induction decay
fMRI	functional magnetic resonance imaging
FWHM	full-width-at-half-maximum
GA	genetic algorithm
GCV	generalized cross-validation
GLM	general linear model
IRR	irregular words
LFP	local field potential
LL	local linearization
MRI	magnetic resonance imaging
MUA	multi-unit activity
NMR	nuclear magnetic resonance
NW	non-words
OSL	Ordinary Least Square
PH	pseudohomophones
rCBF	regional cerebral blood flow
REG	regular words
RF	radiofrequency
SA	simulated annealing
SEM	structural equation modeling

Chapter 1

Introduction

Functional magnetic resonance imaging (fMRI) is a non-invasive technique which provides a powerful tool for understanding the organization of the human brain at the systems level by relating behavioral measures (*e.g.* perceptual, motor or cognitive stimuli or clinical symptoms) to the microscopic alternations in the magnetic field associated with neural activity. The functional MRI response is a complex one depending on intricate physical and physiological phenomena. The advantages of fMRI include its non-invasive nature and relatively high spatial resolution on the order of millimeters with access to submillimeter resolution. In addition, a typical MRI temporal resolution of about 1 s, which can potentially reach the order of 100 ms, is adequate for tracing many specific sequences of neural dynamics. Its ability to cover the entire brain in a period of a few seconds allows researchers to map transient aspects of brain activity. Also, a strength of fMRI is in its flexibility in the design of experimental paradigms, including the possibility of repeated studies of individual subjects and patients.

1.1 Physiological aspects of BOLD contrast

Several distinct parameters can characterize MRI image contrast. First of all, the MRI signal depends on spin density (number of nuclear signals per unit volume), or more specifically, taking into account the primary role of ^1H, on the proton spin density (water concentration). Other contrast relevant parameters include spin-lattice relaxation time T_1, which determine the lifetime of the recovery of the net longitudinal (along the main magnetic field) magnetization after the radiofrequency (RF) pulse, and spin-spin relaxation time T_2, which describes the rate of decay of the net transverse magnetization. Inevitable magnetic field inhomogeneities result in the decay of the signal at a more rapid rate than that due to T_2 relaxation alone. The resultant relaxation is referred to as T_2^* relaxation. It is important to differentiate between coherence loss due to relaxation, which is inherently random and irreversible, and that caused by magnetic field imperfections, which can be reversed. The fMRI signal is determined by precession of the transverse magnetization, the net magnetization in the plane perpendicular to the main magnetic field. The total transverse magnetization M_T is given by

$$M_T(t) = M_0(1 - \exp^{-T_R/T_1})\exp^{-T_E/T_2} \tag{1.1}$$

where repetition time T_R and echo time T are parameters of a pulse sequence used in an experiment. In almost all of the fMRI pulse sequences protocols, a decrease in the T_1 value of the water ^1H signal from a given volume of the brain or an increase in the T_2 value results in a positive MR response.

Prior to the discovery of nuclear magnetic resonance, Pauling and Coryell (1936) described the basic chemical and physical properties of the hemoglobin molecule, the iron-containing oxygen-transport metalloprotein, incorporated in red blood cells, which provide the principal means of delivering oxygen to tissues via the blood. They found that oxygenated hemoglobin is diamagnetic, while deoxygenated hemoglobin is paramagnetic. Presence of paramagnetic hemoglobin affects the local magnetic susceptibility, creating microscopic distortions in the magnetic field within and around blood vessels. This microscopic inhomogeneous magnetic field causes spin dephasing, resulting in a decay of transverse magnetization through T_2^* relaxation. A corollary of this phenomenon is that, theoretically, an MR pulse sequence sensitive to T_2^* contrast should show differences in the decay of transverse magnetization, depending on the proportion of oxygenated hemoglobin in blood (the level of blood oxygenation). This was verified experimentally by Thulborn et al. (1982) with *in vitro* samples. In a seminal paper, Ogawa et al. (1990) demonstrated, *in vivo*, that T_2^*-contrast is sensitive enough to reveal blood oxygenation level dependent (BOLD) signal changes. A similar effect was reported by Turner et al. (1991), who monitored changes in brain oxygenation during periods of recovery from anoxia (deprivation of oxygen supply) and demonstrated that blood behaves as an endogenous susceptibility contrast agent. The BOLD phenomenon was associated not only with macroscopic vessels but was also extended to the brain tissue. It was known that increased neuronal activity is accompanied by increases in both regional blood flow and oxygen consumption, although the stimulus-induced augmentation of cerebral blood flow exceeds the concomitant tissue metabolic rate of oxygen consumption (Fox and Raichle, 1986; Fox

et al., 1988). These observations provided the grounds for the idea that BOLD-based MRI techniques might potentially be useful for exploring brain activation through changes induced by tissue oxygenation. A new era for functional neuroimaging began when three groups (Kwong et al., 1992; Ogawa et al., 1992; Bandettini et al., 1992) reported the success of BOLD-based magnetic resonance imaging of the human brain.

What influences the concentration of oxygenated hemoglobin in the blood, and consequently, the BOLD signal in reaction to neural activity? Studies showed that the BOLD signal depends on contributions from baseline and changes in cerebral blood volume (CBV) (Belliveau et al., 1991), from baseline and changes in cerebral blood flow (CBF) (Kim, 1995), as well as from changes in the cerebral metabolic rate of oxygen consumption $CMRO_2$ (Bandettini et al., 1992). Many studies demonstrated the coupling between neural activity and blood flow (for a review see Raichle (2000)). Increases in glucose utilization, the predominant energy mechanism in the brain, were observed to be approximately proportional to increases in blow flow (Fox and Raichle, 1986; Fox et al., 1988). However, changes in the blood flow and glucose consumption were disproportionately higher than changes in oxygen consumption, suggesting the existence of other metabolic mechanisms, in addition to glycolysis (the metabolic process by which a glucose molecule is oxidized to two molecules of pyruvic acid). Recent experimental measurements have estimated that relative changes in blood flow are two or three times larger than those in oxygen consumption (Davis et al., 1998; Marrett and Gjedde). As a result of this disproportion, the oxygen extraction fraction, and consequently, the deoxyhemoglobin concentration in

blood decreases with activation. Some studies were able to observe an initial dip in vascular oxygenation lasting 1-2 seconds before the standard BOLD increase (Menon et al., 1995). This effect, small and not always present, implies a rapid increase of $CMRO_2$ before the CBF increases to increase the vascular oxygenation.

Despite the basic role of CBF in BOLD imaging, the question regarding the nature of the exact mechanisms of neurovascular coupling remains unresolved. A number of studies indicated a growing role for astrocytes, specifically their ability to control vasodilation via changes in Ca^{2+} concentration (Zonta et al., 2003; Takano et al., 2006). From a geometric point of view, astrocytes are ideally positioned to regulate synaptic transmission and neurovascular coupling, receiving inputs from thousands of neurons and making contact with the local vasculature. Prostaglandins, EET (epoxyeicosatrienoic acids), arachidonic acid and potassium ions are all considered possible candidates for mediating communication between the astrocyte end-feet and vascular smooth muscle (see reviews Haydon and Carmignoto (2006); Koehler et al. (2006)). *In vivo* studies support parallel, independent signaling based on EETs, adenosine and nitric oxide produced by stimulation from nitric oxide synthase (Koehler et al., 2006).

Another critical issue, based on unclear mechanisms of neurovascular coupling and the inherently complex vasculature-dependent nature of the BOLD signal, is the type of neural activity reflected in the dynamics of BOLD signals. Comparing responses from single neuron recordings in monkeys and those from human BOLD studies, two studies (Heeger et al., 2000; Rees et al., 2000) demonstrated a positive correlation between behavioral measures, action potential and fMRI signals. How-

ever, the studies differ dramatically in the quantitative assessment of the rate of change of the spike rate: 0.4 and 9 spikes per second per neuron per 1% change of the BOLD signal in the study by Heeger et al. (2000) and Rees et al. (2000), respectively. These studies provided evidence supporting the idea that BOLD signals reflect neuronal firing rate. However the large discrepancy between these two studies appeared to undermine the indirect approach of those studies. A study by Logothesis et al. (2001) with the anaesthetized monkey used simultaneously recorded fMRI and electrophysiological signals to compare multi-unit spiking activity (MUA) and local field potentials (LFPs) to BOLD responses. This study demonstrated that LFPs, associated with dendritic currents averaged over a population of neurons, rather than the spike rate, are a better indicator of BOLD changes. Lauritzen and Gold (2003) explored the relationship between action potential production, synaptic activity and changes in CBF. Their results indicated that action production is not a necessary condition for a concomitant increase in CBF (or the BOLD signal). The authors emphasized the importance of distinguishing between synaptic inhibition and deactivation when CBF and glucose consumption is increased and decreased, respectively. Further, in another study, Mukamel et al. (2005) compared BOLD changes in the auditory cortex of conscious humans to three types of neuronal activity: spiking activity, low-frequency (5-15 Hz), and high-frequency (40-130 Hz) LFPs. The authors were not able to identify the source behind the BOLD response since the spiking activity was highly correlated with the high-frequency LFPs.

1.2 Spatiotemporal aspects of BOLD contrast

Changes in deoxyhemoglobin concentration in blood is a key component of changes in the BOLD signal. The presence of paramagnetic deoxyhemoglobin has several susceptibility-related effects on the magnitude of the MR signal in general, and on BOLD signal changes, in particular. The deoxyhemoglobin-caused effects may be partitioned into intravascular and extravascular BOLD effects. One intravascular effect is a direct change of the transverse T_2 relaxation of blood (Wright et al., 1991; Thulborn et al., 1982). Another intravascular effect is related to the difference in the phase of the blood-based spins relative to surrounding tissue, due to an oxygenation-dependent susceptibility between the two compartments (Boxerman et al., 1995a). The basis for the extravascular component of BOLD signal is in the main magnetic field inhomogeneities around blood vessels, which can be differentiated between single, straight veins (Yablonskiy and Haacke, 1994) and randomly oriented capillaries (Yablonskiy and Haacke, 1994; Boxerman et al., 1995b).

The spatial properties of vasculature determines the absolute lower limit for the spatial scale that can be measured using BOLD. Specifically, the size of vasculature is about 10 microns in the capillary bed where BOLD changes are well colocalized with the loci of neural activity. In contrast, signal changes can be detected several centimeters away from the large draining veins where the size of vasculature is of the order of a few millimeters (Kim, 1995). Kim et al. (2004) directly compared changes in BOLD signal to single and multi-unit neuronal activity from a large area in the cat brain. Their results revealed a linear correlation between BOLD

changes and neuronal activity at the millimeter scale. However, the correlation between the two measurements varied significantly at the level of individual electrode, leading to the conclusion that BOLD contrast is a robust indicator of neuronal activity only at supra-millimeter scale. Disbrow et al. (2000) used BOLD fMRI and electrophysiological maps to compare the topographic organization of cortical fields. The largest mismatch between the two modalities was found in regions close to larger vessels. In addition, their study indicated that the accuracy or resolution can be significantly degraded in the dimension perpendicular to major local vasculature.

In addition to the issue of spatial localization of neural activity, the evolution (shape) of the BOLD response *per se* may be of separate interest. A typical hemodynamic response might display an initial dip, reaches its maximum within a few seconds after the beginning of a stimulus, and then decaying after the termination of the stimulus with an undershoot in the time course (Ogawa et al., 1998). Cerebral blood flow and BOLD responses are observed to be delayed by a few seconds relative to even a short stimulus (Bandettini et al., 1992). According to Mandeville et al. (1999), the increase of CBF after stimulus onset and decrease after stimulus offset were described with an exponential time constant of 2.4 ± 0.8 seconds. For a prolonged stimulus of about 20 seconds, one can usually observe a plateau, possibly slightly inclined and distorted by an overshoot at the beginning or at the end of the plateau (Kruger et al., 2001). Recent studies indicate that the BOLD response is nonlinear with respect to stimulus duration. Specifically, signal response to shorter duration stimuli were found to be larger than expected from a linear system (Birn et al., 2001). Another aspect of BOLD nonlinearity is a refractory response in which

the signal change is lower in the presence of a preceding identical stimulus. It was reported that the net response to two similar stimuli presented at a run was less than twice the response of a single stimulus alone (Huettel and McCarthy, 2001).

The sampling rate possible for the BOLD signal is limited by the T_1 relaxation. Usually, depending upon the experiment, a typical sampling rate of images acquired with a spatial resolution of $3-4\,\mathrm{mm}$ ranges from $500\,\mathrm{ms}$ to $3-4\,\mathrm{s}$. In a number of studies (Menon et al., 1998; Hernandez et al., 2002) the temporal resolution reached one hundred milliseconds. Within the limits allowed by BOLD physiology, a higher temporal sampling rate can be balanced by a decrease in spatial resolution or through a trade-off with quality of the measured MR signal (Vlaardingerbroek and Den Boer, 1996).

1.3 Structure of fMRI data

Functional MRI data files consist of time series of image volumes. In turn, one image volume is composed of a set of brain slices (images). The images are arrays of greyscale (brightness) values. Each location in an image array is known as a pixel ("picture element") so a pixel represents the spatial location of the MRI signal, represented by the greyscale value, from that location. When the dataset is considered as a volume, the pixels in each slice are known as voxels for "volume elements". Picking up the greyscale values for the same pixel (in other words, for the same voxel) in the same slice (in other words, in the same image) from each volume, we can construct fMRI time series for that voxel. For example, a typical MRI image can

be described by a 128×128 matrix. The number of slices in a typical experiment might be equal to 12. Image acquisition procedure for each slice can be repeated, let us say, 100 times , *i.e.* we have time series with 100 observations. So, the complete data set from such an experiment contains the $128 \times 128 \times 12 \times 100$ values of the fMRI signal.

1.4 Scope of the thesis

In this chapter we very briefly reviewed the physical and physiological basis for blood oxygen level dependent (BOLD) functional magnetic resonance imaging and discussed its intrinsic limits.

A key concern of fMRI-based analysis is that the BOLD signal does not measure neural activity directly. Recently proposed mathematical forward models describe the transduction of neural activity into the BOLD response based on principles and experiments outlined above. In Chapter 2 we transform the expanded "balloon" model into an optimal control problem with the objective of solving the hemodynamic inverse problem, *i.e.* of estimating the dynamics of the neural activity underlying the BOLD signal. We found that the dynamics of the neural signals and the physiological variables can be robustly reconstructed from the BOLD responses. Furthermore, we showed that off/on dynamics of the neural activity is the natural mathematical solution of the model.

The study described in Chapter 3 is an application of the methods proposed in Chapter 2. Specifically, by example of a functional network model from cognitive

neuroscience designed to assess the BOLD and neural activity involvement in a word object interaction decision task, we illustrated how one can recover chronometric information from fMRI time series through solving the hemodynamic inverse problem. In contrast to the hemodynamic level, at the neuronal level, we found statistically significant delays in activation onset between the visual word/object system and the interaction-semantics system, as well as between the visual word/object system and the speech production system.

In general, Chapter 3 considers the problem of smoothing fMRI time courses at the neural level, accompanied with the idea of signal averaging. The same problem was considered in Chapter 4, but at the level of BOLD fMRI time series *per se*. Chapter 4 probes the issue of what amount of smoothing should be considered optimal for fMRI signals, since strong smoothing washes out temporal characteristics. The hemodynamic responses, if not smoothed enough, may reflect irrelevant information. The purpose of the work presented in Chapter 4 was to develop techniques which can estimate the "true" shape of hemodynamic responses and facilitate calculation of BOLD temporal characteristics. Therefore, in Chapter 4 we applied functional data analysis (FDA) techniques based on B-spline smoothing with Tikhonov regularization. In addition, we refined criteria for regularization parameter selection.

Evaluating the temporal parameters of the hemodynamic response implicitly assumes the involvement of smoothing to eliminate variance due to noise. Chapter 5 is an application of the smoothing techniques presented in Chapter 4. The purpose of this part of the thesis was to verify previously reported correlations between the hemodynamic response and the behavioral measures, such as accuracy and reaction

time, extending the applicability of neuroimaging techniques in the issue of integrating cognitive science with neurophysiology. In Chapter 5, we computed maps of several temporal characteristics of BOLD responses measured in Broca's area during a word identification task. We found that BOLD response width is uniquely related to reaction time in the task of naming pseudohomophones. In addition, our results support, in terms of BOLD parameters, the independence between lexical and sub-lexical routes for converting speech to print.

The thesis adopts the manuscript-based format. Chapters 2, 3, 4 and 5 are based on the papers submitted to journals *Neuroimage*, *Statistics in Medicine* and *Brain* (see Preliminaries to each chapter). The bibliography is given after each chapter.

Bibliography

P.A. Bandettini, E.C. Wong, A. Jesmanowicz, R.S. Hinks, and J.S. Hyde. Spin-echo and gradient-echo EPI of human brain activation using BOLD contrast: a comparative study at 1.5 T. *NMR Biomed*, 7(1-2):12–20, 1992.

J.W. Belliveau, D.N.Jr. Kennedy, R.C. McKinstry, B.R. Buchbinder, R.M. Weisskoff, M.S. Cohen, J.M. Vevea, T.J. Brady, and B.R. Rosen. Functional mapping of the human visual cortex by magnetic resonance imaging. *Science*, 254(5032):716–9, 1991.

R.M. Birn, Z.S. Saad, and P.A. Bandettini. Spatial heterogeneity of the nonlinear dynamics in the FMRI BOLD response. *Neuroimage.*, 14(4):817–26, 2001.

F. Bloch. Nuclear induction. *Phys Rev*, 70(7-8):460 – 474, 1946.

J.L. Boxerman, P.A. Bandettini, K.K. Kwong, J.R. Baker, T.L. Davis, B.R. Rosen, and R.M. Weisskoff. The intravascular contribution to fMRI signal change: Monte Carlo modeling and diffusion-weighted studies in vivo. *Magn Reson Med*, 34(1): 4–10, 1995a.

J.L. Boxerman, L.M. Hamberg, B.R. Rosen, and R.M. Weisskoff. MR contrast due to intravascular magnetic susceptibility perturbations. *Magn Reson Med*, 34(4): 555–66, 1995b.

T.L. Davis, K.K. Kwong, R.M. Weisskoff, and B.R. Rosen. Calibrated functional MRI: mapping the dynamics of oxidative metabolism. *Proc Natl Acad Sci U S A*, 95(4):1834–9, 1998.

E.A. Disbrow, D.A. Slutsky, T.P. Roberts, and L.A. Krubitzer. Functional MRI at 1.5 tesla: a comparison of the blood oxygenation level-dependent signal and electrophysiology. *Proc Natl Acad Sci U S A*, 97(17):9718–23, 2000.

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 U S A*, 83(4):11401144, 1986.

P.T. Fox, M.E. Raichle, M.A. Mintun, and C. Dence. Nonoxidative glucose consumption during focal physiologic neural activity. *Science*, 241(4864):462–464, 1988.

A.N. Garraway, P.K. Grannell, and P. Mansfield. Image formation in nmr by a selective irradiative process. *J. Phys. C*, 7:L457–L462, 1974.

P.G. Haydon and G. Carmignoto. Astrocyte control of synaptic transmission and neurovascular coupling. *Physiol Rev*, 86(6):1009–31, 2006.

D.J. Heeger, A.C. Huk, W.S. Geisler, and D.J. Albrecht. Spikes versus BOLD: what does the neuroimaging tell us about neural activity. *Nat. Neurosci.*, 3(7):631–3, 2000.

L. Hernandez, D. Badre, D. Noll, and J. Jonides. Temporal sensitivity of event-related fMRI. *Neuroimage*, 17(2):1018–26, 2002.

S.A. Huettel and G. McCarthy. Regional differences in the refractory period of the hemodynamic response: an event-related fMRI study. *Neuroimage*, 14(5):967–76, 2001.

D.S. Kim, I. Ronen, C. Olman, S.G. Kim, K. Ugurbil, and L.J. Toth. Spatial relationship between neuronal activity and BOLD functional MRI. *Neuroimage*, 21(3):876–85, 2004.

S.G. Kim. Quantification of relative cerebral blood flow change by flow-sensitive alternating inversion recovery (FAIR) technique: application to functional mapping. *Magn Reson Med*, 34(3):293–301, 1995.

R.C. Koehler, D. Gebremedhin, and D.R. Harder. Role of astrocytes in cerebrovascular regulation. *J Appl Physiol*, 100(1):307–17, 2006.

G. Kruger, A. Kastrup, and G.H. Glover. Neuroimaging at 1.5 T and 3.0 T: comparison of oxygenation-sensitive magnetic resonance imaging. *Magn Reson Med*, 45(4):595–604, 2001.

K.K. Kwong, J.W. Belliveau, D.A. Chesler, I.E. Goldberg, R.M. Weisskoff, B.P. Poncelet, D.N. Kennedy, B.E. Hoppel, M.S. Cohen, and R. Turner. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc Natl Acad Sci U S A*, 89(12):5675–9, 1992.

L.D. Landau and L.M Lifshitz. *Quantum mechanics: non-relativistic theory*. Elsevier Science Ltd, 1958.

M. Lauritzen and L. Gold. Brain function and neurophysiological correlates of signals used in functional neuroimaging. *J Neurosci*, 23(10):3972–80, 2003.

P.C. Lauterbur. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242(5394):190 – 191, 1973.

N.K. Logothesis, H. Merkle, M. Augath, T. Trinath, and K. Ugurbil. Neurophisiological investigation of the basis of the fMRI signal. *Nature*, 6843(412):150–157, 2001.

J.B. Mandeville, J.J. Marota, C. Ayata, G. Zaharchuk, M.A. Moskowitz, B.R. Rosen, and R.M. Weisskoff. Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. *J Cereb Blood Flow Metab*, 19(6):679–89, 1999.

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

P. Mansfield and P.K. Grannell. Nmr 'diffraction' in solids? *J. Phys. C*, 6:L422–L426, 1973.

S. Marrett and A. Gjedde. Changes of blood flow and oxygen consumption in visual cortex of living humans. *1997*, 413:205–8.

R.S. Menon, S. Ogawa, X. Hu, J.P. Strupp, P. Anderson, and K. Ugurbil. 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):453–9, 1995.

R.S. Menon, D.C. Luknowsky, and J.S. Gati. Mental chronometry using latency-resolved functional MRI. *Proc Natl Acad Sci U S A*, 95(18):10902–7, 1998.

R. Mukamel, H. Gelbard, A. Arieli, U. Hasson, I. Fried, and R. Malach. Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science*, 309(5736):951–4, 2005.

S Ogawa, T.M. Lee, A.R. Kay, and D.W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S A*, 87(24): 9868–72, 1990.

S. Ogawa, D.W. Tank, R. Menon, J.M. Ellermann, S.G. Kim, H. Merkle, and K. Ugurbil. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci U S A*, 89 (13):5951–5, 1992.

S. Ogawa, R.S. Menon, S.G. Kim, and K. Ugurbil. On the characteristics of functional magnetic resonance imaging of the brain. *Annu Rev Biophys Biomol Struct*, 27:447–74, 1998.

L. Pauling and C.D. Coryell. The magnetic properties and structure of hemoglobin, and carbonmonooxygenated hemoglobin. *Proc Nat Acad Sci U S A*, 22(4):210–36, 1936.

M.E. Raichle. *Brain mapping: the systems*, chapter A brief history of human functional mapping, pages 33–75. San Diego:Academic, 2000.

G. Rees, K. Friston, and C. Koch. A direct quantative relationship between the functional properties of human and macaque V5. *Nat. Neurosci.*, 3(7):719–23, 2000.

T. Takano, G.F. Tian, W. Peng, N. Lou, W. Libionka, X. Han, and M. Nedergaard. Astrocyte-mediated control of cerebral blood flow. *Nat Neurosci*, 9(2):260–7, 2006.

K.R. Thulborn, J.C. Waterton, P.M. Matthews, and G.K. Radda. Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field. *Biochim Biophys Acta*, 714(2):265–70, 1982.

R. Turner, D. Le Bihan, C.T. Moonen, D. Despres, and J. Frank. Echo-planar time course MRI of cat brain oxygenation changes. *Magn Reson Med*, 22(1):159–66, 1991.

M.T. Vlaardingerbroek and J.A. Den Boer. *Magnetic Resonance Imaging: Theory and Practice*. Springer, 1996.

G.A. Wright, B.S. Hu, and A. Macovski. Estimating oxygen saturation of blood in vivo with MR imaging at 1.5 T. *J Magn Reson Imaging*, 1(3):275–83, 1991.

D.A. Yablonskiy and E.M. Haacke. Theory of NMR signal behavior in magnetically inhomogeneous tissues: the static dephasing regime. *Magn Reson Med*, 32(6): 749–63, 1994.

M. Zonta, M.C. Angulo, S. Gobbo, B. Rosengarten, K.A. Hossmann, T. Pozzan, and G. Carmignoto. Neuron-to-astrocyte signaling is central to the dynamic control of brain microcirculation. *Nat Neurosci*, 6(1):43–50, 2003.

Chapter 2

Inferring neural activity from BOLD signals through nonlinear optimization

2.1 Preliminaries

This chapter is based on the manuscript titled "Inferring neural activity from BOLD signals through nonlinear optimization" by Vakorin, V.A., Krakovska, O.O, Borowsky, R. and Sarty, G.E., submitted to *Neuroimage*. The basis for this work is a quantitative link between blood oxygen level dependent (BOLD) fMRI signals and underlying neural activity. The problem is that the BOLD fMRI signal does not measure neuronal activity directly. This fact is a key concern for interpreting functional imaging data based on BOLD. Mathematical models describing the path from neural activity to the BOLD response allow us to numerically solve the inverse problem of estimating the timing and amplitude of the neuronal activity underlying the BOLD signal. In fact, these models can be viewed as an advanced substitute for the impulse response function.

In this work, the issue of estimating the dynamics of neuronal activity from the observed BOLD signal is considered within the framework of optimization problems. The model is based on the extended "balloon" model and describes the conversion of

neuronal signals into the BOLD response through the transitional dynamics of the blood flow-inducing signal, cerebral blood flow, cerebral blood volume and deoxyhemoglobin concentration. Global optimization techniques are applied to find a control input (the neuronal activity and/or the biophysical parameters in the model) that causes the system to follow an admissible solution to minimize discrepancy between model and experimental data. As an alternative to a local linearization (LL) filtering scheme, the optimization method escapes the linearization of the transition system and provides a possibility to search for the global optimum, avoiding spurious local minima.

We have found that the dynamics of the neural signals and the physiological variables as well as the biophysical parameters can be robustly reconstructed from the BOLD responses. Furthermore, it is shown that spiking off/on dynamics of the neural activity is the natural mathematical solution of the model. Incorporating, in addition, the expansion of the neural input by smooth basis functions, representing a low-pass filtering, allows us to model local field potential (LFP) solutions instead of spiking solutions.

2.2 Introduction

Functional magnetic resonance imaging (fMRI) is a non-invasive technique used to gain insights into brain function. The most common contrast available for researchers is blood oxygen level dependent (BOLD) contrast (Ogawa et al., 1990) based on the dependence of the magnetic properties of hemoglobin on its oxygenation state in

reaction to evoked neuronal activity. Unfortunately BOLD fMRI does not directly measure neuronal electrical activity nor does it straightforwardly reflect the changes in brain energy metabolism. Geometrically, the BOLD contrast depends both on the intravascular (Boxerman et al., 1995a,b; Hoogenraad et al., 2001) and extravascular (Ogawa et al., 1993; Hoogenraad et al., 2001) dephasing effects caused by changes in deoxyhemoglobin concentration. As indicated in a study addressing the spatial correspondence between BOLD and neuronal activity, separation of the effects due to intravascular and extravascular components is not observed, which complicates the issue of localizing the neuronal activity underlying BOLD responses (Kim et al., 2004). Further, many studies have demonstrated the quite complex nature of the BOLD signal as a function of cerebral blood flow, blood volume and oxygen metabolism (for reviews see Heeger et al. (2000); Heeger and Ress (2002); Logothesis et al. (2001); Logothetis (2002); Nair (2005); Sarty (2007)). From a temporal point of view, the dynamics of BOLD are delayed relative to electrical activity because of slow vascular transitional effects. These spatiotemporal discrepancies between the correspondence of the BOLD responses with the underlying neuronal activity may undermine researchers' confidence in the validity of results obtained using BOLD fMRI for many cognitive and perceptual studies.

Being in fact an echo of neuronal activity, the BOLD effect blurs temporally evolving events when separated episodes of activity are filtered by the temporal buffer of the hemodynamics. Therefore, the desire to put more meaningful interpretation on fMRI time courses is a key motive for transferring the stress of the analysis from the BOLD level down to neuronal level. In other words, there is a need to solve the

problem of extracting the relevant neuronal information from the measured hemodynamic response, which can be called the inverse hemodynamic problem (Buckner, 2003). Combined neuroimaging studies and electrophysiological recordings (Logothesis et al., 2001; Heeger et al., 2000; Rees et al., 2000; Mukamel et al., 2005) have demonstrated a coupling between BOLD and neuronal activity, laying the basis to believe that considering the inverse hemodynamic problem makes good sense.

At first, indirect comparison between responses from single neuron recordings in monkeys and human fMRI measurements (Heeger et al., 2000; Rees et al., 2000) demonstrated a positive correlation between behavioral measures, action potential and fMRI signals. Specifically, the fMRI response in human cortical areas V1 and V5 was found to be proportional, although not strictly linearly, to the firing rate of individual neurons in the same areas in monkeys. Assuming that these brain areas are homologous and that firing rates in two species are similar, the authors were able to estimate the average firing rate per neuron corresponding to the changes in fMRI signal.

Further, simultaneously recorded fMRI signals and neuronal potentials were compared in a study with the anaesthetized monkey (Logothesis et al., 2001) to evaluate the linear model. The neuronal signals observed as spikes superimposed on waves of lower frequencies were measured with a low impedance microelectrode placed in the extracellular space in V1. A high-pass filter (300-1500 Hz) was used to obtain multi-unit activity (MUA) which is believed to reflect spiking activity around the electrode tip. Applying a low-pass filter (40-130 Hz) resulted in local field potentials (LFP). The basis for LFPs, the low-frequency component of the extracellular field potentials,

is thought to be a synchronized dendritic current averaged over a neural population. The study addressed the issue of predicting the fMRI measurements from the LFP and MUA. It was estimated that the LFPs accounted for approximately 7.6% more of the variance in the fMRI signals than did the MUA. The difference was found to be statistically significant. Nevertheless, the authors did not exclude the possibility that spikes are a key determinant of the BOLD response despite the slightly better predictive power of the LFPs (Logothetis and Wandell, 2004).

The same linear convolution model was used to examine the coupling between fMRI and neuronal activity in the auditory cortex of conscious humans (Mukamel et al., 2005). Three types of neuronal activity were tested: spiking activity, low-frequency (5-15 Hz) LFPs, and high-frequency (40-130 Hz) LFPs. All three potential predictors were shown to be significantly correlated with the fMRI response. The spiking activity in Heschl's gyrus of one patient was found to be a statistically significant, slightly better predictor of the BOLD signal than the high-frequency LFPs, while for another patient there was no significant difference in predictive power. Since the spiking activity was highly correlated with the high-frequency LFPs, the authors were not able to identify the source behind the BOLD response. However the findings support, at least under a natural stimulus condition, the view that BOLD signals are a measure of the average firing rate of the underlying population.

The invasive nature of electrophysiological methods dramatically limits their application for studying brain function. However, convolution techniques, motivated by the need for modeling interactions at a neural level, were introduced in pure fMRI analysis before neurovascular coupling was demonstrated. The problem of inferring

the dynamics of the neuronal activity from BOLD can be identified with the deconvolution problem which implicitly assumes the existence of the intrinsic dynamic system for converting underlying neuronal response into observed BOLD signal. In the simplest case, neuronal activity based on the stimulus paradigm can be convolved (smoothed) with the impulse response function approximated by a Poisson distribution function (Friston et al., 1994). Later on, manipulations with the stimulus amplitude and duration showed the nonlinear characteristics of the BOLD response (Vazquez and Noll, 1998).

To account for non-linear effects, the convolution method was extended using Volterra series, an expansion of a dynamic, nonlinear functional in analogy to the Taylor series (Friston et al., 1998). The Volterra series can be considered as time-invariant high-order convolution kernels applied to a stimulus in order to simulate the observed BOLD signal. The authors considered the particular case of a known input sequence that varies from voxel to voxel only in terms of the scaling parameters. As a result, estimations of the neuronal activity function obtained by using this approach strongly depend on *a priori* knowledge regarding the experimental design and parametrization of the Volterra kernels. In this approach, the quantification of activation patterns that are not under direct experimental control, may be distorted. Specifically, this can be an issue for the voxels from those brain areas that are indirectly time-locked to the stimulus in an experiment designed for studying dynamic interactions through the estimation of the effective connectivity.

Within the framework of Volterra kernels, an expectation-maximization algorithm (an iterative procedure) was used to estimate the parameters of the hemo-

dynamic response (Friston et al., 1998). These search methods converge relatively quickly but there exists a significant disadvantage in their tendency to converge to a local optimum depending on starting values. Further, a key concern for convolution, linear or non-linear, is the variability of the observed hemodynamic responses which are reported to be different not only across subjects but also from voxel to voxel as well as from task to task (Handwerker et al., 2004). The implication is that the underlying neuronal activity is not approximated well by the higher order convolution of fixed functions with stimulus timing waveforms.

Biophysical models of neurovascular coupling (Friston et al., 2000; Zheng et al., 2002) allow one to model this coupling and use it as new information (essentially, *a priori* information) in the effort to infer the transitional dynamics. The models can be considered as an advanced substitute for the impulse response function approach. These forward space-state models produce a predicted BOLD response for a combination of hemodynamic parameters and an input function representing the neuronal dynamics. The "balloon" model proposed by Buxton et al. (1998) and the windkessel model of Mandeville et al. (1999) describe the coupled dynamics of cerebral blood volume (CBV), deoxyhemoglobin concentration and their effect on the BOLD signal as a function of cerebral blood flow (CBF) into an elastic vascular system (balloon). The models are based on the idea of mechanical expansion of the venous compartment, induced by increased CBF. The balloon model was extended by Friston et al. (2000) by implementing the dynamics of a vasoactive agent concentration, a flow-inducing signal defined in terms of the rate of change of normalized blood flow. The outcome is a continuous case of a state-space model consisting of

a set of nonlinear differential equations (process equations) describing the dynamics of CBF, the CBF-inducing signal, CBV, deoxyhemoglobin concentration and an observation (measurement) equation to define the BOLD signal. The flow-inducing signal was assumed to be self-diffusive and generated by the underlying neuronal activity. However the interpretation of the neuronal activity function remains unclear as the extended balloon model does not reflect the kinetics of the neuronal signalling mechanisms controlling the blood flow.

In a previous study, a local linearization (LL) filter was proposed (Riera et al., 2004) to solve the hemodynamic inverse problem formulated as a space state model. The LL method belongs to a class of recursive filters reduced to linearized equations of evolution for the conditional means of the state variables and their covariance matrix. It can be used to reconstruct the underlying neuronal activity together with the dynamics of state variables and some of the biophysical parameters. Although the LL filter is thought to overcome the problem of stability typical for recursive filters, it may suffer from erroneous estimations of the dynamics of the neuronal activity underlying the BOLD signal, as may be suggested from the visual analysis of the results reported (Riera et al., 2004).

As an alternative to the LL method, we propose a new formalism to tackle the issue of extracting the dynamics of the neural activity from the measured BOLD signal when the mathematical model is considered within the framework of optimization problems. The space-state model is transformed into an optimal control problem. Specifically, the problem under consideration is to find an admissible control or input (the neural activity and/or the biophysical parameters in the model) that causes the

system to follow an admissible solution that minimizes a performance measure based on model-experiment discrepancy. We aim to (i) clarify the meaning of the neuronal activity function used as the input for the intrinsic dynamic system that converts the neuronal activity into BOLD and (ii) introduce a robust method to solve the hemodynamic inverse problem.

2.3 Framing the problem and methods of solution.

2.3.1 Optimal control problem. Pontryagin minimum principle.

The expanded balloon model (Friston et al., 2000) has been previously introduced to link the neural activity $u(t)$ to the BOLD signal $h(t)$ on the time interval $[0, T]$, with time t measured in seconds. The coupling between the neuronal input and the evoked BOLD signal is specified with a space-state model as depicted in Fig. 2.1. Specifically,

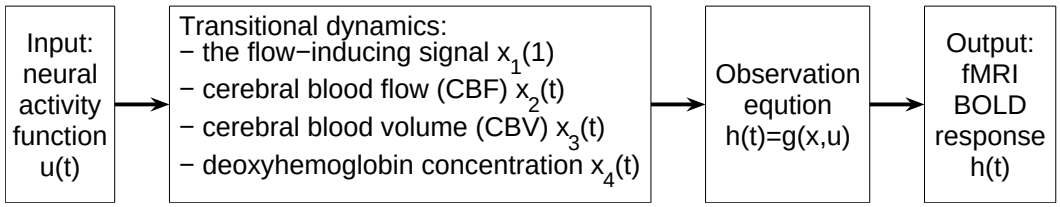


Figure 2.1: A diagram describing the expanded balloon model. The intrinsic biophysical system that determines the transition of the neuronal signal to the BOLD response is formulated in terms of a state-space model. The input is the neuronal activity function. It defines the intermediate dynamics of 4 physiological variables, namely a flow-inducing signal, CBF, CBV and deoxyhemoglobin concentration. The functions for CBV and deoxyhemoglobin concentration are, in turn, used in the observation equation that produces the BOLD response.

the first part describes the transitional dynamics of the state-space vector $\boldsymbol{x}(t) = (x_1(t), x_2(t), x_3(t), x_4(t))^T$, where the functions given by $\boldsymbol{x}(t)$ represent

$$\begin{aligned} x_1(t) &= \text{the flow-inducing signal} \\ x_2(t) &= \text{cerebral blood flow (CBF)} \\ x_3(t) &= \text{cerebral blood volume (CBV)} \\ x_4(t) &= \text{deoxyhemoglobin concentration} \end{aligned}$$

where all quantities are normalized relative to resting values.

The second part of the model is the observational equation for the BOLD signal $h(t)$. The system is defined as follows:

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \mathbf{f}(\boldsymbol{x}(t), u(t)) \tag{2.1}$$

$$h(t) = g(\boldsymbol{x}(t), u(t)) \tag{2.2}$$

the explicit form of which is

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = f_1(\boldsymbol{x}(t), u(t)) = u(t) - \frac{x_1(t)}{\tau_s} - \frac{x_2(t) - 1}{\tau_f} \tag{2.3}$$

$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = f_2(\boldsymbol{x}(t), u(t)) = x_1(t) \tag{2.4}$$

$$\frac{\mathrm{d}x_3}{\mathrm{d}t} = f_3(\boldsymbol{x}(t), u(t)) = \frac{1}{\tau_0}\left(x_2(t) - x_3(t)^{\frac{1}{\alpha}}\right) \tag{2.5}$$

$$\frac{\mathrm{d}x_4}{\mathrm{d}t} = f_4(\boldsymbol{x}(t), u(t)) = \frac{1}{\tau_0}\left(\frac{x_2(t)}{E_0}\left[1 - (1 - E_0)^{\frac{1}{x_2(t)}}\right] - x_4(t)x_3(t)^{\frac{1-\alpha}{\alpha}}\right) \tag{2.6}$$

with initial conditions

$$\boldsymbol{x}^\top(0) = (x_1(0), x_2(0), x_3(0), x_4(0)) = (0, 1, 1, 1) \tag{2.7}$$

and

$$h(t) = g\left(\boldsymbol{x}(t), u(t)\right) = V_0\left(k_1(1 - x_4(t)) + k_2\left(\frac{1 - x_4(t)}{x_3(t)}\right) + k_3(1 - x_3(t))\right) \tag{2.8}$$

where $\Theta \equiv \{\tau_s, \tau_f, \tau_0, \alpha, E_0, V_0, k_1, k_2, k_3\}$ is the set of biophysical parameters.

The changes in blood flow determine the dynamics of the blood volume and deoxyhemoglobin content. In particular, an increase in deoxyhemoglobin content $x_4(t)$ results in an decreased BOLD signal $h(t)$. The BOLD response depends on changes in concentration of the deoxyhemoglobin, and therefore relies on a balance between the delivery and consumption of oxygen, on the one hand, and on the cerebral blood volume, on the other hand.

According to the model (Buxton et al., 1998), the increased blood flow inflates the balloon, the venous compartment, which dilutes the deoxygenated blood. Another assumption is the capacity of the swelled venous balloon to expel deoxygenated blood at a greater rate (this models the relationship between the blood outflow and the blood volume). Changes in blood volume and clearance of deoxyhemoglobin determine the output, the BOLD signal. The BOLD response which is a non-linear function of deoxyhemoglobin and blood volume, is divided into extravenous and intravenous components with weights representing relative volumes. The rate of change of deoxyhemoglobin is defined by the difference between deoxyhemoglobin entering the venous compartment and that expelled from the compartment. The rate of change of blood volume is the blood inflow minus the flood outflow. The rate of changes of the blood flow is defined in terms of the concentration of a vasoactive agent released by neural activity. In this model, the concentration of the vasoactive agent is denoted the blood flow-inducing signal. This signal is generated by neuronal activity. The dynamics of the flow-inducing signal is affected by two mechanisms. First, the model includes a mechanism for signal decay (elimination). Second, there

is a autoregulatory feedback on the dynamics of the signal from the blood flow.

The model depends on a set of constant parameters. These include the time constant of the flow-inducing signal decay τ_s, the time constant of the feedback regulatory mechanism τ_f, the mean transit time of a blood cell in the venous compartment τ_0, Grubb's exponent α for the flow-volume relation, baseline oxygen extraction rate E_0, and resting blood volume fraction V_0. While E_0, α, V_0, k_1, k_2 and k_3 are dimensionless, the parameters τ_s, τ_f and τ_0 are measured in units of time. Specifically, τ_0 represents the time it takes to traverse the venous compartment. The time constant τ_s can be associated with the half-life of vasoactive agents that control vasodilation (see Parri and Crunelli (2003) on the issue of possible candidate mechanism for neurovascular coupling). The parameter τ_f reflects the effective influence time of the autoregulatory feedback on the vasoactive signal from the blood flow. The values of the constant parameters Θ can be estimated directly or indirectly (Friston et al., 2000). The parameters k_1, k_2 and k_3 depend on experimental and physiological parameters and reflect the contribution effects of CBF and deoxyhemoglobin concentration to BOLD signal. It has been previously estimated for 1.5 T and T_E=40 ms that $k_1 \simeq 7E_0$, $k_2 \simeq 2$ and $k_3 \simeq 2E_0 - 0.2$ (Boxerman et al., 1995a; Ogawa et al., 1993).

As a note here, the BOLD signal is an echo of neural activity. It implies that there exists a temporal buffer between the hemodynamic response and neural activity. Three time constants, namely τ_0, τ_s and τ_f, reflect the temporal characteristics of this buffer and model the delay between the neural activity and the evoked hemodynamic response.

Assuming that the neuronal activity $u(t)$ determines the dynamics of the physiological variables $\mathbf{x}$ and thus the shape of the BOLD response given in (2.8), we may infer the dynamics of the neural activity function. A criterion for choosing $u(t)$ is to minimize a discrepancy between the model and experiment, or putting it mathematically, a distance $D(\cdot,\cdot)$ between two functions, namely the observed BOLD $\hat{h}(t)$ and the function $g\left(\boldsymbol{x}(t), u(t)\right)$ calculated through the transition dynamics (2.3)-(2.8) for a given $u(t)$. So the objective is to find

$$\hat{u}(t) = \arg\min_{u(t)} D(h(u), \hat{h}) \tag{2.9}$$

where the distance between two functions is defined in the L^2 sense as follows

$$D(h, \hat{h}) \equiv D(u) \equiv \int_0^T \Big(\hat{h}(t) - g\left(\boldsymbol{x}(t), u(t)\right) \Big)^2 \mathrm{d}t \tag{2.10}$$

We assume that the variable to be optimized is the function $u(t)$ with values normalized to be between 0 and 1, where $u(t) = 1$ represents a totally activated voxel and $u(t) = 0$ represents no activation. We choose for U the control class for the functions $u(t)$

$$U \equiv \{u(t) : \ u \text{ is Lebesgue-measurable with values between 0 and 1}\} \tag{2.11}$$

The problem of minimizing the objective function given in (2.9) subject to the system (2.3)-(2.6) which governs the transitional dynamics of the vector $\boldsymbol{x}(t)$ with the initial conditions given in (2.7) is a typical problem in optimal control theory. In other words, the problem under consideration is to find an admissible control $u(t)$ that causes the system (2.3)-(2.6) to follow an admissible trajectory $\boldsymbol{x}^T(t) =$

$(x_1(t), x_2(t), x_3(t), x_4(t))$ that minimizes the performance measure given by (2.10) on an interval with fixed final time T.

We invoke the Pontryagin minimum principle (Pontryagin, 1962) to determine the optimal control law and the necessary properties of the optimal $u(t)$. The principle states that the Hamiltonian $\mathcal{H}(\boldsymbol{x}(t), u(t), \boldsymbol{p}(t))$, a function describing the state of a system in terms of state variables $\boldsymbol{x}(t)$ and costate variables $\boldsymbol{p}(t)$, must be minimized over the set of all the possible control functions $u(t)$. The Hamiltonian of the system is defined as

$$\mathcal{H}(\boldsymbol{x}(t), u(t), \boldsymbol{p}(t)) = \big(\hat{h}(t) - g(\boldsymbol{x}(t))\big)^2 + \sum_{j=1}^{4} p_j(t) f_j(\boldsymbol{x}(t)) \tag{2.12}$$

where the functions $\boldsymbol{p}^T = \big(p_1(t), p_2(t), p_3(t), p_4(t)\big)$ represent the Lagrange multipliers (the adjoint functions) for the derivatives $\mathbf{f}^T = \big(f_1(\mathbf{x}), f_2(\mathbf{x}),\ f_3(\mathbf{x}),\ f_4(\mathbf{x})\big)$, respectively. The first term in (2.12) is a minimum square criterion, while the second term constrains $\boldsymbol{x}(t)$ to be an admissible solution of the system (2.3)-(2.6). The Pontryagin principle states necessary conditions for u^*, $\boldsymbol{x}^*$ and $\boldsymbol{p}^*$ to be optimal for all $t \in [0, T]$ in terms of the Hamiltonian and its partial derivatives. Specifically, if the pair $(u^*(t), \boldsymbol{x}^*(t))$ is optimal for the problem (2.9) and (2.3)-(2.6), there must exist functions $\boldsymbol{p}^*(t)$ such that the following conditions hold:

$$\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}t} = \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}^*}\big(\boldsymbol{x}^*(t), u^*(t), \boldsymbol{p}^*(t)\big) \tag{2.13}$$

$$\frac{\mathrm{d}\boldsymbol{p}^*}{\mathrm{d}t} = -\frac{\partial \mathcal{H}}{\partial \boldsymbol{x}^*}\big(\boldsymbol{x}^*(t), u^*(t), \boldsymbol{p}^*(t)\big) \tag{2.14}$$

$$\mathcal{H}(\boldsymbol{x}^*(t), u^*(t), \boldsymbol{p}^*(t)) = \min_{v} \mathcal{H}(\boldsymbol{x}^*(t), v, \boldsymbol{p}^*(t)) \tag{2.15}$$

The equations in (2.13) are simply the state system (2.3)-(2.6). The other set of differential equations given in (2.14) describes the behavior of the adjoint system. It

represents the transitional dynamics of the Lagrange multipliers $\boldsymbol{p}(t)$ through a set of four non-linear differential equations. This component should be considered together with a set of transversality conditions, which determines the boundary conditions for the adjoint variables $\boldsymbol{p}(t)$. In the case with a fixed final time T, the transversality conditions are reduced to the condition

$$\boldsymbol{p}^T(T) = \big(p_1(T), p_2(T), p_3(T), p_4(T)\big) = (0, 0, 0, 0) \tag{2.16}$$

which is to be used along with (2.7). Finally, condition (2.15) states that the function $u(t)$ is to be selected such that $u(t)$ minimizes the Hamiltonian.

The function $u^*(t)$ is a control input that causes the Hamiltonian $\mathcal{H}(\boldsymbol{x}^*(t), u(t), \boldsymbol{p}^*(t))$ to reach its global minimum. However, it should be noted that the system (2.13)-(2.14) with the boundary conditions (2.7) and (2.16) accounts only for necessary conditions for optimality. To derive an explicit condition for $u^*(t)$ to be optimal we note that the Hamiltonian (2.12) is a linear function of $u(t)$ with the coefficient of proportionality equal to $p_1(t)$. So, the Hamiltonian $\mathcal{H}(\boldsymbol{x}(t), u(t))$ can be sorted in a way that combines like terms with respect to $u(t)$

$$\mathcal{H}(\boldsymbol{x}(t), u(t), \boldsymbol{p}(t)) = p_1(t)u(t) + \mathcal{R}(\boldsymbol{x}(t), \boldsymbol{p}(t)) \tag{2.17}$$

where $\mathcal{R}(\boldsymbol{x}(t), \boldsymbol{p}(t))$ is a function that does not depend on the function $u(t)$. In this specific case, for $u^*(t)$ to minimize the Hamiltonian, as determined by (2.15), it is necessary that

$$u^*(t) = \begin{cases} 1 & \text{if } p_1^*(t) < 0 \\ 0 & \text{if } p_1^*(t) > 0 \\ \text{any value between 0 and 1 inclusive} & \text{if } p_1^*(t) = 0 \end{cases} \tag{2.18}$$

Condition (2.18) reflects the fact that if $p_1^*(t)$ is negative at some point t, in order to minimize the term $p_1^*(t)u^*(t)$ and in turn the Hamiltonian, the function $u^*(t)$ must take on the highest possible positive value, namely the value of 1 as defined by (2.11). In case of positive $p_1^*(t)$, the expression $p_1^*(t)u^*(t)$ will be minimized with $u^*(t) = 0$. To sum up, the value of the first Lagrange multiplier at any time point determines the values of the function $u(t)$ at this point.

It is quite challenging for the problem defined through (2.13)-(2.14),(2.7),(2.16) and (2.18) to be solved analytically in the terms of simple functions. Numerical techniques are better applied to find a solution. With a numerical approach, the exact solution $u^*(t)$ is to be approximated with a vector $\boldsymbol{u} = (u_1, \dots u_i \dots u_n) = (u(t_1), \dots u(t_i) \dots u(t_n))$ considered at sample points t_i with finite resolution $\triangle t = t_i - t_{i-1} = \frac{T}{n-1}$. According to the theoretical result (2.18) derived from applying the Pontryagin principle, the optimal $u(t_i)$ at any point t_i should take on values of only zero or unity, excluding numerically unfeasible cases of $p_1^*(t) = 0$. Such on/off dynamics of the neuronal activity function $u(t)$ can be interpreted as a spike train on the interval $[0, T]$. In other words, spiking dynamics is the natural mathematical solution for the hemodynamic inverse problem based on the expanded balloon model considered here. Also, it is worth mentioning that the piece-wise step function, traditionally used to represent a stimulus design, is a specific case of the function (2.18).

2.3.2 Multidimensional optimization problem.

As the result of applying the Pontryagin principle, a system of 8 differential equations, the state and costate equations (2.13)-(2.14), and 1 algebraic relation (2.18) must be satisfied on the interval $[0, T]$. However, there is one factor that prevents one from simply solving the differential system by numerical integration. The boundary conditions are split: for the state variables they are defined at the initial point of the interval while the boundary conditions for the adjoint variables are specified at the final point. The boundary conditions at the beginning of the interval do not define a unique solution on their own. Due to the existence of many solutions that satisfy the starting boundary conditions, a selected solution may not satisfy the boundary conditions at the end. In fact, an iteration procedure is necessary to unite the separated boundary conditions to form a single global solution.

The shooting method and relaxation methods are examples of the numerical techniques available to solve the two-point boundary problem (Keller, 1968). Nevertheless, this problem can be avoided, and the optimal $u^*(t)$ can be found without explicitly solving the system (2.13)-(2.14) and (2.18). The optimal control problem defined in (2.9) can be considered from the perspective of a multidimensional optimization scheme. From this point of view, the objective is to minimize the functional

$$D(\boldsymbol{u}) \equiv D(h(\boldsymbol{u}), \hat{h}) \equiv \int_0^T \Big(\hat{h}(t) - g\left(\boldsymbol{x}(t), \boldsymbol{u}\right) \Big)^2 \mathrm{d}t \tag{2.19}$$

in the problem

$$\hat{\boldsymbol{u}} = \arg \min_{\boldsymbol{u}} D(\boldsymbol{u}) \tag{2.20}$$

over an unknown vector $\mathbf{u} = (u_1, \ldots, u_n)$ in the n-dimensional space. The important thing to note here is that regardless of the numerical method applied to solve the minimization problem (2.20), the separate theoretical result obtained in (2.18) from the Pontryagin minimum principle remains valid, and this information can be used in implementing algorithms to numerically solve the minimization problem.

The evaluated on/off dynamics of the solution can be associated with the multi-unit spiking activity (MUA), a high-frequency component of the extracellular field potentials. The relationship between BOLD fMRI signals and the spike activity as the neural basis of the BOLD has been addressed in a number of studies (Heeger et al., 2000; Logothetis, 2002; Mukamel et al., 2005; Rees et al., 2000). In some of those studies, local field potentials (LFPs) have been implicated as a more preferable index of neural activity for BOLD (Logothetis, 2002). Mathematically, LFPs represent the low-frequency range of extracellular field potentials.

One way of implementing a low-pass filter in the optimization problem is to model a smoothed representation of the input $u(t)$ by a linear combination of known basis functions, $\phi_j(t)$ as follows

$$u(t) = \sum_{j=1}^{K} c_j \, \phi_j \tag{2.21}$$

where the K-vector $\boldsymbol{c} = \{c_j\}_{j=1}^{K}$ represents the coefficients c_j. To test the proposed techniques, we used a B-spline basis (de Boor, 1978). B-splines were chosen because of their stable numerical properties, continuous derivatives at the joining points and local support which prevents a single observation from affecting the entire shape of the hemodynamic response (Carnicer and Pena, 1994; Ramsay and Silverman, 2002).

Furthermore, estimating the physiological parameters $\boldsymbol{\theta}$ in the set Θ which influence the dynamics of the transitional system may be of some interest. Including those parameters $\boldsymbol{\theta}$ into the optimization scheme, the problem defined in (2.9) becomes

$$\{\boldsymbol{c}, \boldsymbol{\theta}\} = \arg \min_{\boldsymbol{c},\boldsymbol{\theta}} \int_0^T \Big(\hat{h}(t) - g\left(\boldsymbol{x}(t), \boldsymbol{\theta}, \boldsymbol{c}\right) \Big)^2 \mathrm{d}t \tag{2.22}$$

where both the coefficients of the basis function expansion $\boldsymbol{c}$ and the biophysical parameters of interest $\boldsymbol{\theta}$ are to be selected in order to minimize the discrepancy between the model and experiment.

We note that the problem defined in (2.3)-(2.6) and (2.22) is a multi-dimensional minimization problem over the space spanned by the vectors $\boldsymbol{\theta}$ and $\boldsymbol{c}$. The solution to this problem provides estimates both of the coefficients $\mathbf{c} = \{c_j\}$ of the neural function expansion and the physiological parameters $\boldsymbol{\theta}$ under consideration. Consequently, the dynamics of the neuronal activity underlying the observed BOLD response $\hat{h}(t)$ can be evaluated in the span of the basis $\phi_i(t)$ through (2.21).

2.4 Implementation

2.4.1 Numerical techniques.

The proposed methods were tested according to the scheme illustrated in Fig. 2.2. First, we assumed a specific neuronal activity function $u(t)$. Two types of neural function were designed so as to model both LFP and MUA activity. The stepwise function considered at discrete points with finite time resolution represented the MUA activity, while in the case of LFPs, some of the coefficients c_j of the approx-

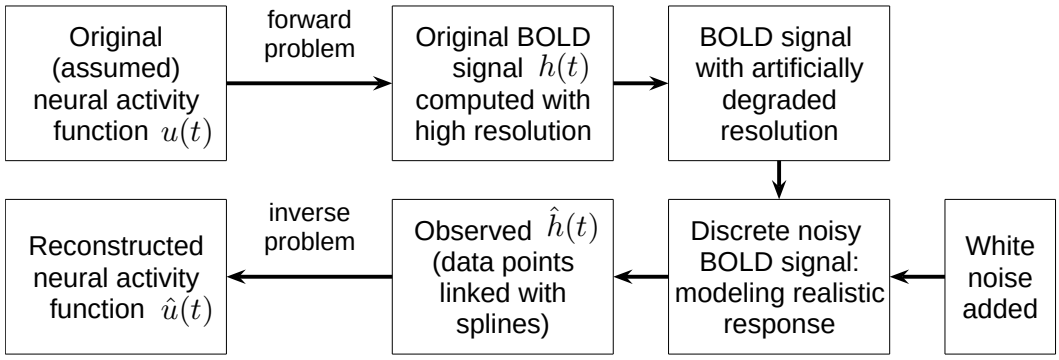


Figure 2.2: A schematic diagram of the step-by-step procedure used to test the techniques for solving the hemodynamic inverse model. First, we solved the forward model based on an assumed neuronal activity function. Then, the quality of the modeled BOLD signal was degraded. Finally, the inverse problem was solved to produce the reconstructed neuronal function which was compared with the original one.

imation (2.21) were set equal to one, with the others equal to zero. Then, given the assumed (original) neural activity function $u(t)$, the forward problem (2.3)-(2.8) was solved to produce the corresponding BOLD signal. Next, the resolution of the BOLD response related to the original neural activity was decreased. In addition, white noise was added to the modeled BOLD signal in order to make it more realistic. At this step, the noisy discrete BOLD function was assumed to model the observed fMRI signal evoked by the assumed neural activity. Finally, the discrete points of the BOLD signal were linked together by splines in order to improve the resolution of the BOLD signal for the goal to solve the set of the differential equations. This newly smoothed function containing low-frequency noise was used to solve the inverse hemodynamic problem.

The core of the scheme designed to implement the proposed methods is in the numerical techniques used for solving the hemodynamic inverse problem. In the previous section, we discussed two approaches for numerically estimating the neural

activity underlying the observed BOLD. The first approach was concerned with solving the system of optimality conditions (2.13)-(2.14),(2.7),(2.16) and (2.18) obtained from applying the Pontryagin minimum principle. However it was noted that this is a two-point boundary value problem wherein boundary conditions (2.7) and (2.16) are separated at times $t = 0$ and $t = T$. It turns out that the system (2.13)-(2.14) is hard to solve numerically because of the unstable behavior of the adjoint variables $\boldsymbol{p}(t)$. So, another way of minimizing the discrepancy between the model and experiment is to consider the problem of minimizing the functional (2.9) from the perspective of a multi-dimensional optimization problem.

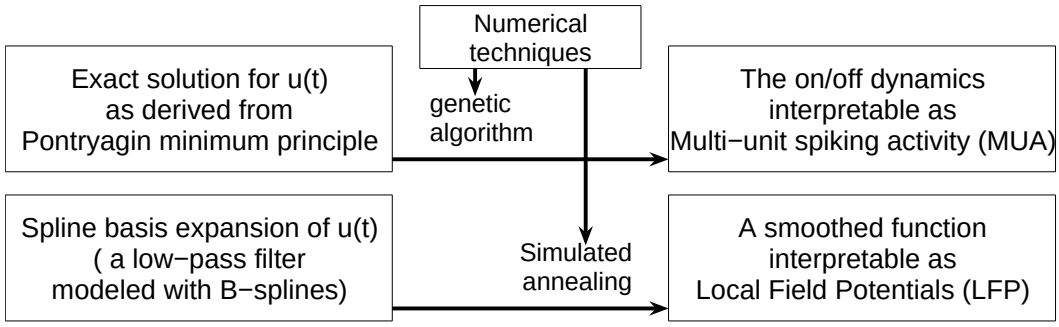


Figure 2.3: Two numeric techniques, genetic algorithm and simulated annealing, were used to solve the multidimensional minimization problems. In the context of multi-unit activity, a genetic algorithm was used to reconstruct the natural solution, the on/off neuronal function as predicted from Pontryagin minimum principle. The smoothed neuronal function and the parameters of the biophysical model were reconstructed by simulated annealing in the context of LFPs.

Two numerical techniques, namely a genetic algorithm (GA) and simulated annealing (SA), were used to solve the minimization problems (2.20) and (2.22) as illustrated in Fig. 2.3. GA was used to minimize (2.20) and to estimate the neural activity function assumed to represent the MUA activity. The information regarding the on/off dynamics of the function to be estimated, derived from the Pontryagin

principle, can be used in the minimization scheme as it determines the general shape of the neural function. A genetic algorithm (for review, see Glover and Kochenberger (2003)) is a search technique from the class of numerical methods targeted to look for global solutions to optimization problems. GAs are a particular class of evolutionary algorithms that mimic the mechanisms of evolutionary biology including inheritance, mutation, natural selection and crossover. Traditionally, solutions are represented in binary as strings of 0s and 1s, but different encodings are also possible (Michalewicz, 1996; Potter et al., 1998). In each generation, the cost function associated with each string in the whole population is evaluated, the best strings are selected from the current population, which are then recombined and mutated to form a new population. The procedure is repeated in the next iteration. In the case of the minimization problem (2.20), the neural activity $u(t)$ is approximated by the binary string $\boldsymbol{u} = (u_1, \ldots, u_n)$ with a resolution $(t_{i+1} - t_i)$ representing the spike train to be optimized.

In the case of a neural function represented by LFPs, simulated annealing (for a review, see Glover and Kochenberger (2003)) was used to numerically search for the optimal solution in the minimization problem over the space spanned by the vector $\boldsymbol{c}$ and the parameters $\boldsymbol{\theta}$. SA is a local search method capable of escaping from local minima by using limited hill-climbing steps, moves which increase the objective function value. It is based on the idea of taking a random walk at successively lower temperatures, where the probability of taking a step up is given by a Boltzmann distribution. At each iteration of the algorithm, the objective function generates a value that falls into one of two possible cases. A step will occur if the new value

of the functional is lower. If the new value is higher, the transition can still occur, and its likelihood is proportional to the temperature T and inversely proportional to the difference between the old and new values of the objective function. In the case of the minimization problem (2.22), the objective function is minimized over the space spanned by the coefficients of the linear expansion (2.21) and the set of the biophysical parameters $\boldsymbol{\theta}$. To test the proposed techniques, the vector $\boldsymbol{\theta}$ included the system parameters τ_s, τ_f and τ_0, *i.e.* $\boldsymbol{\theta}^T = (\tau_s\, \tau_f\, \tau_0)$.

A few points are worth mentioning regarding the GA and SA. First, using GA or SA for minimizing the objective function (2.20) or (2.22) respectively, yields a forward approach for inverting the problem. In contrast with the Pontryagin principle, which is used to convert the optimal control problem into a set of differential equations to provide optimality conditions, the GA and SA do not convert the problem at all, but simply search for solutions in a heuristic and intelligent way.

Second, both GA and SA are capable of escaping local minima. However, these techniques are not always guaranteed to find the global minimum. Moreover, SA and GA will not always converge to a set of solutions associated with the global attractor. In general, the question of how and why genetic algorithms work remains open (Reeves and Rowe, 2001). Convergence results and criteria for simulated annealing applied to continuous global optimization problems are only now being specified (see, for example, Yang (2000) and Locatelli (2000)). Nevertheless, in many cases the converging properties of these numerical methods to provide a reasonable and stable approximation to the global optimum (for a review on GA and SA see Glover and Kochenberger (2003)).

A major disadvantage of simulated annealing algorithms is their slow convergence speed. As indicated in literature on SA (Glover and Kochenberger, 2003), necessary conditions for convergence tend to require slower cooling schemes which, in turn, imply a slow rate of convergence. GAs are generally considered to be faster then SA, however they are still computationally expensive. For the present study, the average computation time for either algorithm was about 10-15 hours per one fMRI time course on a AMD Athlon 2400+ computer with 2GHz CPU and 1.5GB of RAM.

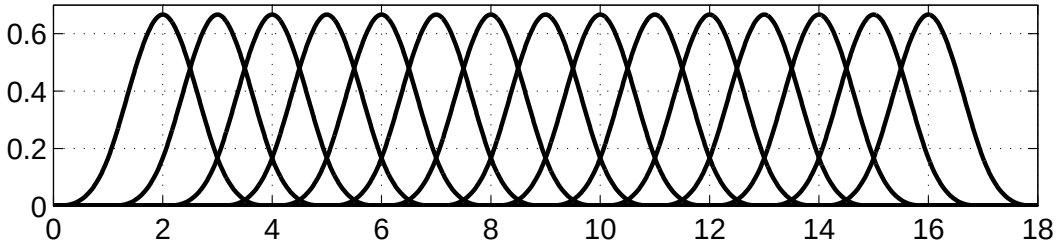


Figure 2.4: A set of B-spline functions. A B-spline is a spline function which has minimal support with respect to a given degree, smoothness, and domain partition. Each basis function $\phi_j(t)$ illustrated is a piecewise polynomial of the third degree.

An order four B-spline system $\{\phi_j\}_{j=1}^{K}$ (de Boor, 1978) was used for the basis function expansion (2.21) of the neural activity function $u(t)$ as shown in Fig 2.4. The knots were placed at the data points t_i where the BOLD signal was assumed to have been observed. A B-spline is a spline function which has minimal support with respect to a given degree, smoothness, and domain partition. Each basis function $\phi_j(t)$ is a spline function defined by an order and a knot sequence. Specifically, for a given $(m+1)$ knots t_i with $t_0 \leq t_1 \leq \ldots \leq t_{m-1} \leq t_m$ a function $u(t)$ can be parameterized with basis B-splines $\phi_{i,n}(t)$ of degree n (or order $n+1$) as

$$u(t) = \sum_{i=0}^{m-n-1} c_i \, \phi_{i,n}(t) \quad \text{for } t \in [t_n, t_{m-n}] \tag{2.23}$$

where the parameters c_i (coefficients of linear expansion in the one-dimensional case) are called control points. The basis (m-n) B-splines can be defined using the de Boor recursion formula

$$\phi_{j,0}(t) = \begin{cases} 1 & \text{if } t_j \leq t \leq t_{j+1} \\ 0 & \text{otherwise} \end{cases} \tag{2.24}$$

$$\phi_{j,n}(t) = \frac{t - t_j}{t_{j+n-t_j}} \phi_{j,n-1}(t) + \frac{t_{j+n+1} - t}{t_{j+n+1} - t_{j+1}} \phi_{j+1,n-1}(t) \tag{2.25}$$

When the knots are equidistant (as in case of fMRI time series with the repetition time equal to T_R), the basis B-splines are just shifted copies of each other.

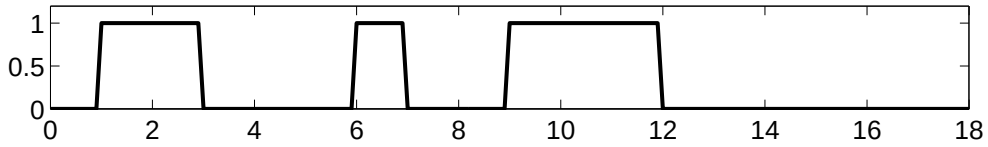
2.4.2 Numerical results.

Fig. 2.5 shows one of the simulation examples for the MUA problem (2.20) solved using the genetic algorithm. The original neural spiking activity that might be initiated by an event-related task (Fig. 2.5(a)) is illustrated in Fig. 2.5(b). The BOLD response evoked by the specified neural activity was calculated through the transition system (2.3)-(2.8). Then, the BOLD response evaluated through the forward approach was mixed with noise. The resulting simulated BOLD signal used to solve the inverse hemodynamic problem is illustrated in Fig. 2.5(d) as a solid line. Finally, Fig. 2.5(c) shows the neural activity function reconstructed with a GA through minimizing the functional of (2.20). The following values of the biophysical parameters were used in the simulations: $\tau_s = 0.8\,\text{s}$, $\tau_f = 0.4\,\text{s}$ and $\tau_0 = 1\,\text{s}$ (Friston et al., 2000); $\alpha = 0.4$ (Grubb et al., 1974; Mandeville et al., 1999); $E_0 = 0.6$ which is within the range of the values of 0.4 and 0.8 as reported in Buxton et al. (1998) and Friston

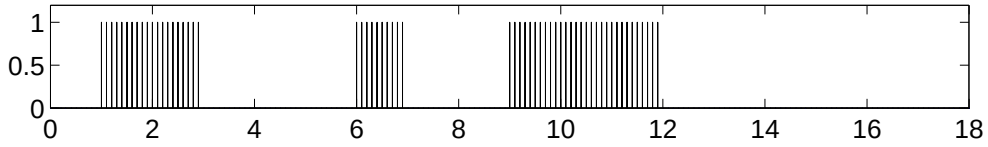
et al. (2000), respectively.

The application of simulated annealing was used to solve the inverse problem (2.22), in the case of neural activity assumed to be represented by LFPs, as demonstrated in Fig. 2.6. The goal was to simultaneously estimate the coefficients $\boldsymbol{c}$ of the neural activity function expansion (2.21) and three biophysical parameters: the time constant of flow-inducing signal decay τ_s, the time constant of the feedback regulatory mechanism τ_f, and the mean transit time in the venous compartment τ_0. The remaining values of the physiological parameters were kept the same as in the GA case discussed above. The "observed" BOLD response (solid line, Fig. 2.6(b)) containing white noise was elicited from the specified neural activity, the dynamics of which is depicted as the solid line in Fig. 2.6(a). The solution to the inverse problem in the sense of (2.22) is superimposed on the original dynamics of neural activity in Fig. 2.6(a). As shown in Fig. 2.6(b), the BOLD signal evoked by the reconstructed neural function closely matches the "observed" BOLD. The estimated values of the unknown physiological parameters were found to be approximately $\tau_s = 0.76\,\mathrm{s}$, $\tau_f = 0.42\,\mathrm{s}$ and $\tau_0 = 1.18\,\mathrm{s}$, while the original were $\tau_s = 0.8\,\mathrm{s}$, $\tau_f = 0.4\,\mathrm{s}$ and $\tau_0 = 1.0\,\mathrm{s}$. Fig. 2.5 and 2.6 show some mismatch between the reconstructed and the original neural signals. There exist two major reasons for this mismatch. First of all, it is inherent in the nature of genetic and simulated annealing algorithms to produce an approximate solution to the optimization problem. The matter is also complicated by the presence of noise added to the "observed" BOLD signals to make them more realistic (see the testing scheme in Fig. 2.2).

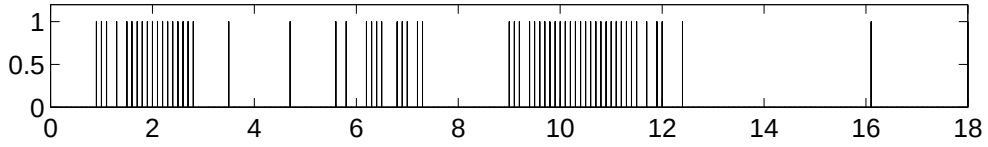
In addition to simulations, data from a functional imaging experiment, conducted



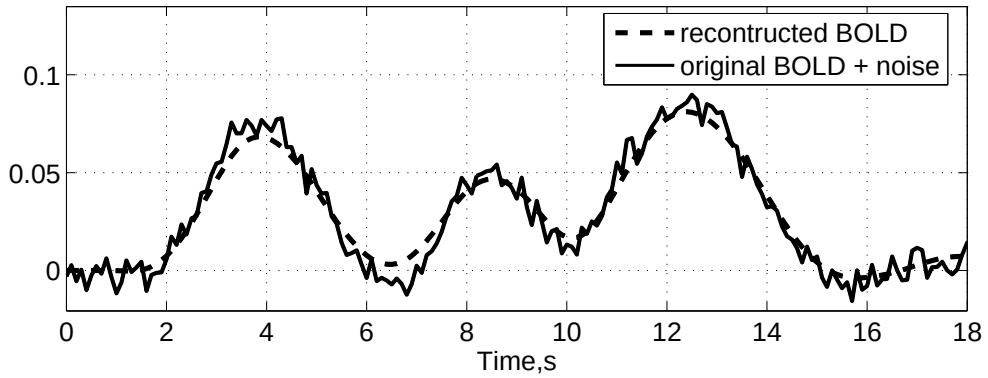
(a) the assumed stimulus timing $u(t)$, a piecewise step function with values between one and zero



(b) the spiking neural activity function associated with the stimulus function in (a), considered at discrete points with a finite resolution

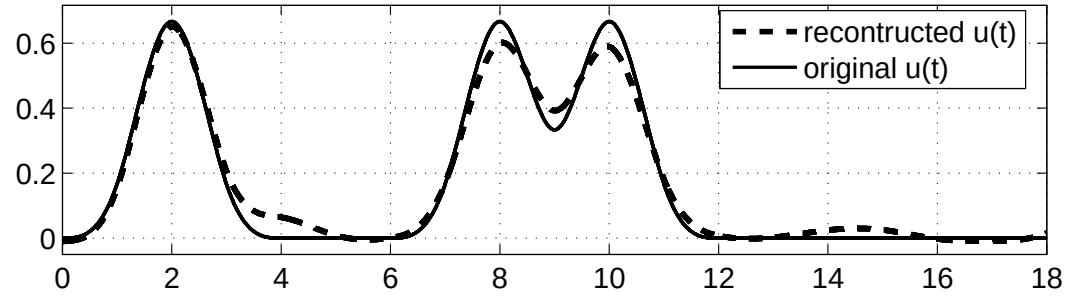


(c) the spike train reconstructed from the "observed" BOLD signal

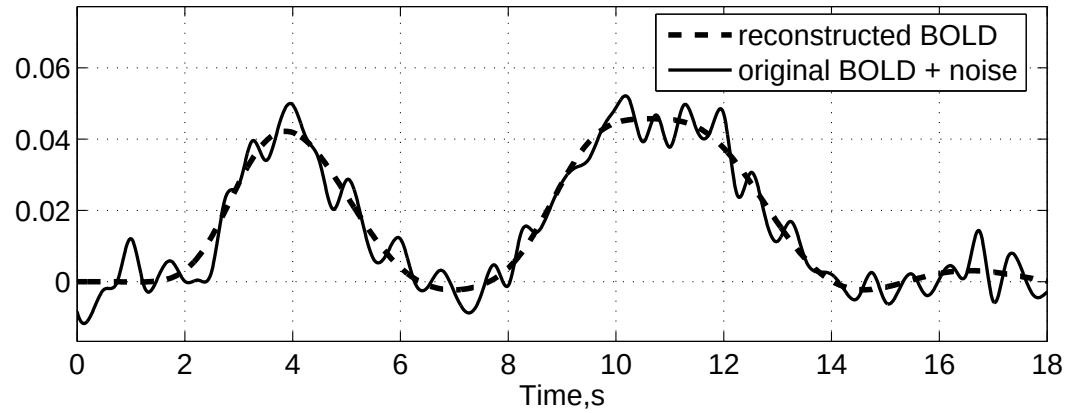


(d) the assumed dynamics of the BOLD signal mixed artificially with white noise and superimposed on the reconstructed BOLD response

Figure 2.5: The dynamics of BOLD signal and the underlying neural activity in the forward-inverse relationships in the case of the MUA simulations. The bars in (b) and (c) correspond to the values of unity, while the absence of a bar at a point stands for the value of zero. The simulated BOLD signal (solid line, (d)) evoked by the specified neural activity (b) was used to solve the inverse problem (2.20) using a genetic algorithm. The solution to the inverse problem is the spiky train (c) underlying the reconstructed BOLD response (dashed line, (d)).

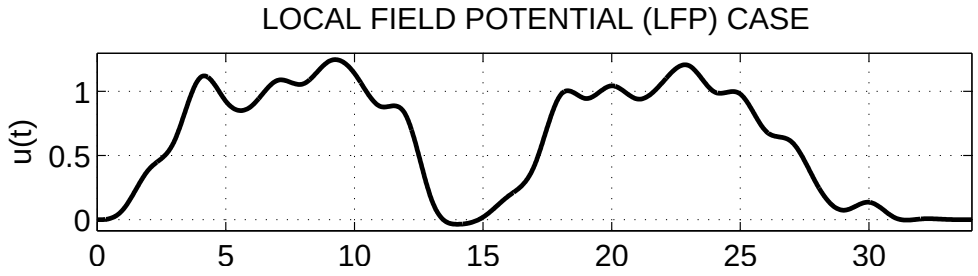


(a) the original smoothed function modeling the neural activity superimposed on the reconstructed dynamics of the neural activity

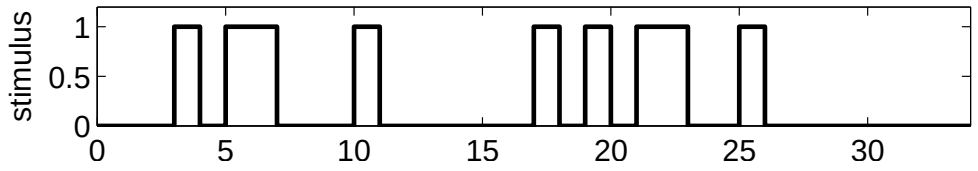


(b) the "observed" BOLD signal superimposed on the evaluated BOLD response evoked by the reconstructed neural activity function

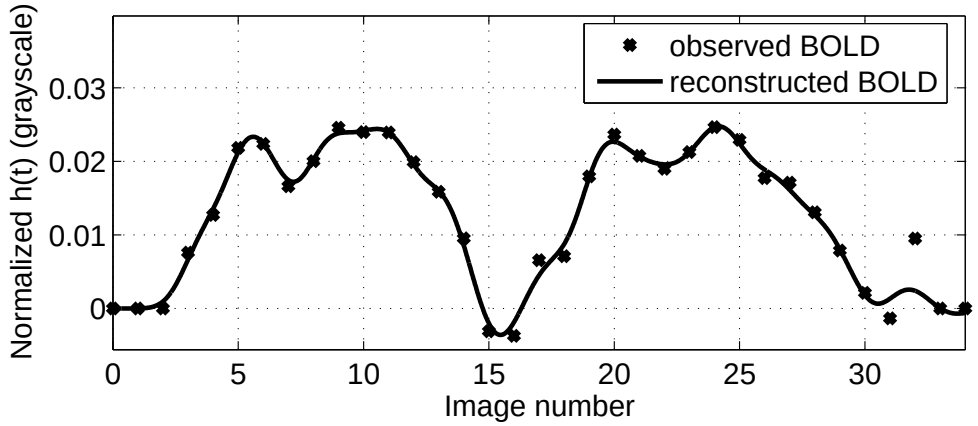
Figure 2.6: The dynamics of BOLD signal and the underlying neural activity in the forward-inverse relationships in the case of the LFP simulations: the original smoothed function (solid line) modeling the neural activity superimposed on the reconstructed dynamics (dashed line) of the neural activity (a); and the "observed" BOLD signal superimposed on the evaluated BOLD response evoked by the reconstructed neural activity function (b). The original (assumed) BOLD response (solid line, (b)), which was mixed with white noise, is associated with the specified neural activity (solid line, (a). The reconstructed neural activity function (dashed line, (a)) was evaluated using simulated annealing by minimizing the discrepancy between the model and experiment in the form of the functional (2.22).



(a) the estimated function $u(t)$ as a combination of the B-splines basis functions



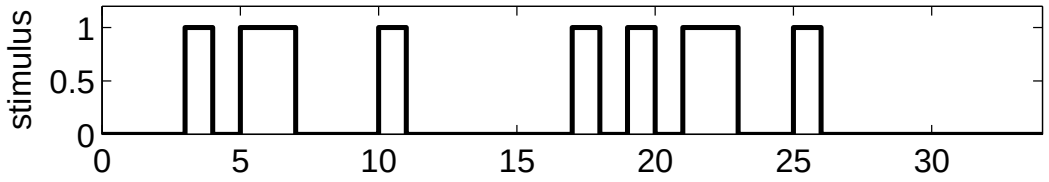
(b) the stimulus timing waveform



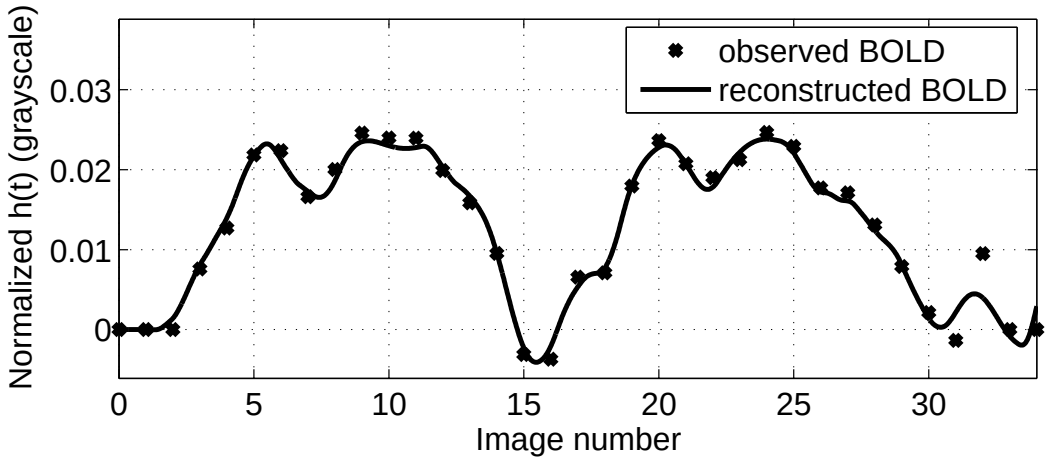
(c) the estimated BOLD dynamics superimposed on the normalized original time course

Figure 2.7: The dynamics of BOLD signal and the underlying neural activity as the simulated annealing solution to the hemodynamic inverse problem for an activated voxel in an experiment with event-related design: the estimated function $u(t)$ as a combination of the B-splines basis functions (a); the stimulus timing waveform (b); and the estimated BOLD dynamics superimposed on the normalized original time course (c).

(a) the estimated function $u(t)$, an on/off function as derived from the Pontryagin minimum principle

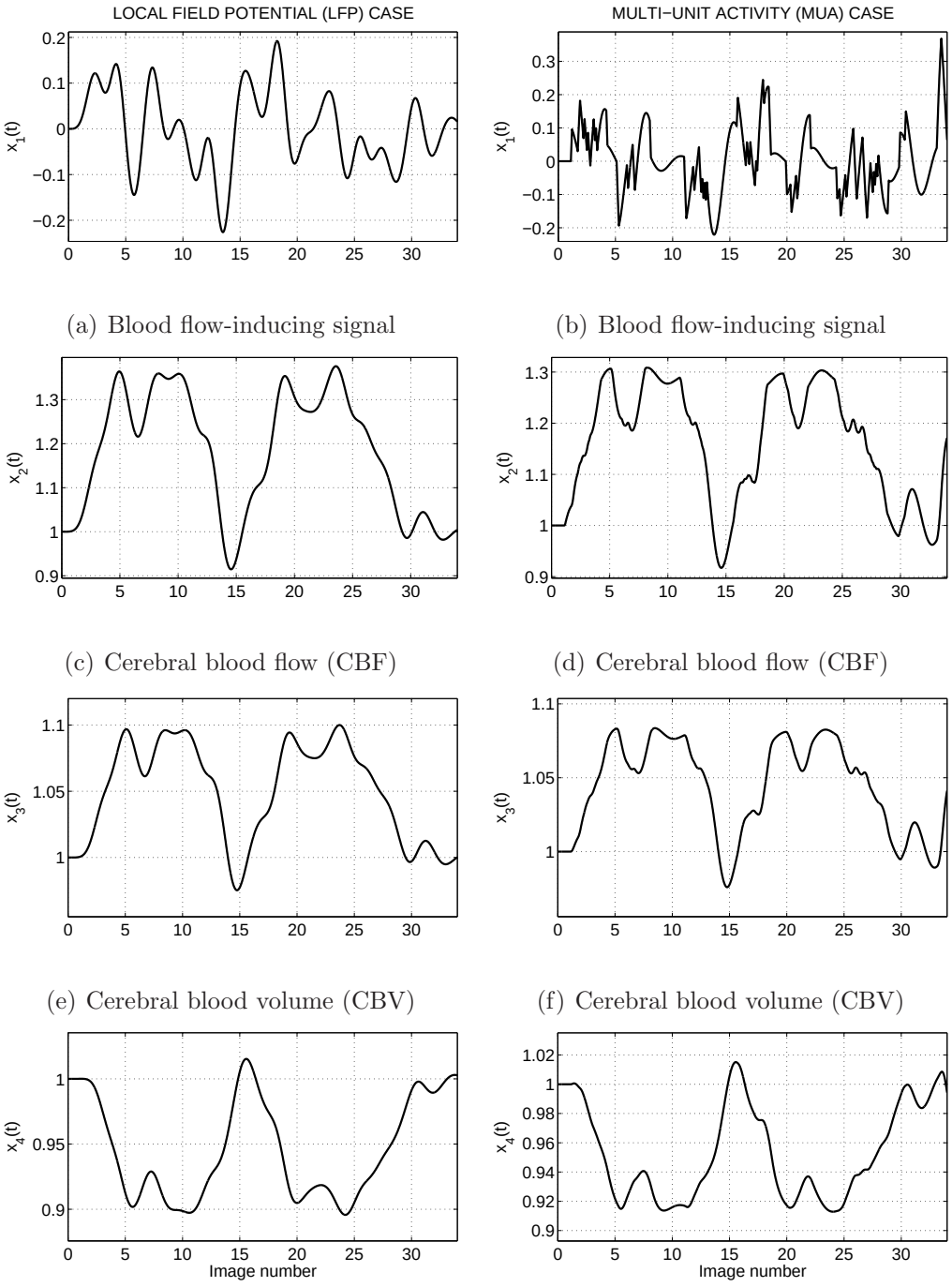


(b) the estimated BOLD dynamics superimposed on the normalized original time course



(c) the estimated BOLD dynamics superimposed on the normalized original time course

Figure 2.8: The dynamics of BOLD signal and the underlying neural activity as the genetic algorithm solution to the hemodynamic inverse problem for the same activated voxel used in Fig. 2.7: the estimated function $u(t)$, an on/off function as derived from the Pontryagin minimum principle (a); and the estimated BOLD dynamics superimposed on the normalized original time course (b); and the estimated BOLD dynamics superimposed on the normalized original time course (c).



(a) Blood flow-inducing signal

(b) Blood flow-inducing signal

(c) Cerebral blood flow (CBF)

(d) Cerebral blood flow (CBF)

(e) Cerebral blood volume (CBV)

(f) Cerebral blood volume (CBV)

(g) Deoxyhemoglobin concentration

(h) Deoxyhemoglobin concentration

Figure 2.9: The transitional dynamics of the four physiological variables $\boldsymbol{x}(t)$ in two cases, LFP(left column) and MUA(right column), associated with the BOLD signals plotted in Fig 2.7(c) (LFP case) and Fig 2.8(c) (MUA case), respectively: the dynamics of the blood flow-inducing signal $x_1(t)$ (a,b); the dynamics of the cerebral blood flow $x_2(t)$ (c,d); the dynamics of the cerebral blood volume $x_3(t)$ (e,f); and the dynamics of the deoxyhemoglobin concentration $x_4(t)$ (g,h).

with a 1.5 T Siemens Symphony Magnetom (Erlanger, Germany) scanner, was used to further test the algorithms. A gradient echo T_2^* single-shot EPI sequence (T_E = 55ms and T_R = 1600ms) was used with fat saturation pulses. The 12 acquired slices had a square field of view of 250 mm and were 8-mm thick with an interslice gap of 2 mm. The original 64×64 data matrices were Fourier reconstructed to 128×128 images. The experiment design was a mixture of event-related and altering block designs. After 3 dummy volumes, 6 blocks of 15 volumes of images were acquired for a total of 93 volumes per run. Of the 15 data volumes per block, the first 9 were regarded as stimulus. Within the first 9 volumes, two tasks were presented in a non-regular interchanging order, while the last 6 volumes were acquired with the subject at rest.

The study we took the fMRI time series from consisted of 6 people (average age about 26). We randomly selected data from one the subject data sets. Next, we randomly selected a voxel from brain region found activated during task performance (in the superior parietal gyrus). To delineate the activated regions, 2 task impulse response waveforms were fitted to the original fMRI data with the AFNI deconvolution algorithm (Cox, 1996). The task was to identify presentations to the right or left visual fields. Specifically, the subject was given a 2 alternative forced choice (2AFC) task in which a colored circle was presented either to the left or right of fixation, and immediately followed by another colored circle in the same location that was either the same color or a different color. The participant responded "same" or "different" by pressing one of two buttons on a mouse that was MRI compatible.

The input stimulus timing (see Fig. 2.7(b)) for each task was periodic with the

period equal to 2 blocks of volumes acquired. The time series of the 3 resulting periods were folded over and averaged with the goal of increasing signal-to-noise ratio, producing the basis for the observed BOLD signal used to solve the hemodynamic inverse problem (2.22). This averaged signal was normalized with respect to the intensity value from the last volume in the sequence used to achieve the steady state of image contrast before the main body of the experiment. To alleviate edge effects, the time span was extended by adding 2 zero values to the normalized BOLD time course before and after the main body. To increase the sampling rate of the BOLD signal for the purpose of numerically integrating the system of differential equations, the discrete adjacent points of the BOLD time series were linked together by linear splines. The neural activity function $u(t)$ was expanded in a cubic B-splines basis (see Fig. 2.4) constructed by using a knot sequence placed at the moments of volume acquisition. The vector of the physiological parameters $\boldsymbol{\theta}$ was exposed to optimization jointly with the coefficients $\boldsymbol{c}$ included τ_s, τ_f and τ_0 as in the case of the LFP simulations discussed above.

Two methods, both simulated annealing and a genetic algorithm were used to solve the inverse hemodynamic problem and to find the underlying neuronal activity function in the LFP and MUA sense, respectively. Figs. 2.7 and 2.9 (left column) represent the LFP example solution demonstrating the relations between the observed BOLD signal, the dynamics of neuronal activity underlying this signal, stimulus timing and the transitional dynamics of the physiological variables $\boldsymbol{x}(t)$ in an activated voxel for the stimulus waveform illustrated in Fig. 2.7(b). The normalized BOLD signal is illustrated in Fig. 2.7(c), while Fig. 2.7(a) shows the estimated dynamics

of the underlying neuronal activity. The physiological parameters $\boldsymbol{\theta}$ were estimated as $\hat{\tau}_s = 1.36\,T_R = 2.2\,\mathrm{s}$, $\hat{\tau}_f = 0.28\,T_R = 0.45\,\mathrm{s}$ and $\hat{\tau}_0 = 0.59\,T_R = 0.94\,\mathrm{s}$. Two of these parameters, specifically τ_s and τ_f are of the same order of magnitude as reported in previous studies (Friston et al., 2000; Riera et al., 2004). However, the value of the last parameter $\hat{\tau}_0 = 0.94\,\mathrm{s}$ is the same as that reported in Friston et al. (2000), but is in contrast to the value found by the local linearization scheme (Riera et al., 2004) which provides much higher values for this parameter. This should not be taken as the final word on the value of physiological parameters expected from this SA technique. The aim here was to demonstrate the applicability of SA, not to take a comprehensive look at the inferred values. Finally, it can be noted that the BOLD response visually fails to capture peculiarities of the stimulus paradigm function, but the estimated neuronal dynamics in both cases, although far from being perfect, somewhat better reflects the structure of the stimulus timing. To a greater degree, this statement is valid for second part of the stimulus design function: clusters of spikes in the MUA case (Figs. 2.8(a)) and neuronal episodes in the LFP case (Fig. 2.7(a)) can be attributed to the stimulus condition of the stimulus timing waveform.

Figs. 2.8 and 2.9 (right column) show the MUA example solution and the transitional dynamics of the physiological variables $\boldsymbol{x}(t)$ for the same voxel which was used to solve the inverse hemodynamic problem with SA. The parameters used in the MUA scheme were set as defined in the LFP case. When visually compared to the performance of simulated annealing (LFP case), the genetic algorithm (MUA case) leads, roughly speaking, to the same degree of approximation of the observed

BOLD with the modeled function. A difference in performance between two cases is the ability of the implemented algorithms to reconstruct the transitional dynamics. As can be seen, the dynamics of deoxyhemoglobin concentration, blood volume and blood flow in the LFP case do not differ much from that in the MUA case. The dynamics of the blood flow-inducing signal is generated by neuronal activity. In the MUA case, the presence of higher frequencies in the flow-inducing signal is explained by a spiky input, while the smoothed input in the LFP case produces the smoothed dynamics of the signal $x_1(t)$. If we look at the dynamics of blood flow in the LFP case, we can see that the high-frequency component of the blood flow is gone. This is explained by the specificity of the equation (2.4): the rate of change of blood flow is proportional to the concentration of a vasoactive agent (the flow-inducing signal) released by neuronal activity. From a mathematical point of view, equation (2.4) is nothing else but a smoothing filter for the signal $x_1(t)$. In fact, this equation is a key element of the model, which reflects the presence of a temporal buffer between neuronal activity and hemodynamic responses. Searching for the optimal $u(t)$ in the basis of B-splines can be considered as an additional component of this buffer. In this context, the shape of the neural signals underlying BOLD responses is defined by the assumptions as to what type of neural activity is best associated with the hemodynamic response.

2.5 Discussion

We developed a technique which allows one to solve the inverse hemodynamic problem, *i.e.* to reconstruct the neuronal dynamics underlying the BOLD responses, based on a previously proposed mathematical forward model. In addition, this approach proved to be useful in the issue of interpreting the meaning of the neuronal function used in the model. To achieve these goals, the space-state model that produces the dynamics of the BOLD signal through the transitional dynamics of flow-inducing signal, CBF, CBV and deoxyhemoglobin concentration was transformed into an optimization problem.

Generally speaking, solving the inverse hemodynamic problem should imply being able to differentiate between various possible types of neuronal activity in any voxel at any time: local field potentials, current source density, some measures of excitatory/inhibitory postsynaptic activity, sub-threshold membrane potential oscillations, or firing patterns of a subpopulation of neurons. It should be clarified that solving the hemodynamic inverse problem in a general sense should not be confused with the methods proposed in the present study. The basic assumption of our approach is the validity of the previously proposed mathematical model of the transduction of neural activity to the BOLD signal. Therefore, if the intrinsic path is specified, solving the hemodynamic problem makes perfect sense.

There are several advantages in considering the problem of inferring neuronal activity from the perspective of optimization. Within the framework of the Pontryagin minimum principle, necessary conditions for optimality can be defined in terms of

a Hamiltonian, a function characterizing the total energy of system as a function of state and costate variables. We can subsequently make general inferences about the behavior of the neuronal function to be estimated.

The biophysical model considered in this work and other similar models (Zheng et al., 2002) do not directly specify the meaning of the neuronal activity function. As found from applying the Pontryagin minimum principle, the natural solution to the optimization problem is a function which may take only its lowest and highest values. Specifically, we were working on the assumption that the neuronal function is normalized to be within the limits of the values of zero and one. The theoretical result regarding the switching nature of the neuronal dynamics was due to linear dependence of the Hamiltonian on the neuronal activity function. Several interpretations for this fact are possible. First, the on/off dynamics of the neuronal activity can be interpreted as a spike train, if it is considered on the microscopic level. In other words, it models an approach wherein multi-unit spiking activity is considered as an indicator of changes in BOLD signal. The distance between spikes is defined by the sampling rate of numerical techniques used to solve the optimization problem. Theoretically, the underlying neuronal dynamics can be approximated with a function considered at points with a sampling rate of very high frequency. Practically, it is limited by the computational power of the numerical algorithms applied. Increasing the number of dimensions would inexorably also decrease the ability of the numerical techniques to converge to the global optimum.

Another interpretation of the on/off function is possible from the macroscopic perspective of design functions. Specifically, a piece wise step paradigm function

representing block design or event-related design can be considered as a particular case of the solution predicted *a priori* from the Pontryagin minimum principle. This may partly justify the approach adopted in Friston (2002) when an expectation-maximization (EM) algorithm was applied to estimate the input function expanded as a linear combination of known functions associated with the stimulus. Furthermore, abstractly decreasing the width of the step function representing a separate event would correspond to going from the macroscopic to the microscopic level. As an extreme case, the step function standing for a single event with decreasing duration would be turned into a spike, which in turn may model a quantal release of neurotransmitters.

In addition, the proposed techniques allow one to categorize the neuronal activity function on the scale of LFPs rather than in terms of MUA. As simultaneous MRI and electrophysiological recordings indicate, LFPs are a slightly better indicator of BOLD changes than MUA (Logothetis, 2002). In the context of LFPs, a positive, conventional pattern of the hemodynamic response is associated with an increase of underlying neuronal activity (Logothetis, 2002), while a negative BOLD signal is related to local decreases in neuronal activity below the on-going level (Shmuel et al., 2006). Being essentially the low-frequency component of electrophysiological signals, LFPs can be modeled through an expansion of the neuronal activity function in a basis of some known functions acting as a low-pass filter.

The Pontryagin minimum principle can be used not only in the issue of interpreting the neuronal input to the models but also can serve as a bridge between two schemes, optimization and integrating a system of equations. Non-linear opti-

mization and solving the non-linear system of equations are closely related processes. However they are not equivalent as some information usually becomes lost during translation. Spurious stationary point solutions to the optimization problem expressed as a non-linear system of equations often do not correspond to local optima of the original problem, and *vice versa*. Nevertheless, in both cases the task is to find the absolutely best set of parameters among the spurious solutions, taking into account the non-linear nature of the problem.

Considering the hemodynamic inverse problem from the optimization point of view paves the way towards straightforwardly applying numerical techniques known to find global optima. Genetic algorithms and simulated annealing are examples of such methods. These techniques are able to escape local optima even when the structure of the searching space is quite complicated. They do not need linearization and, in general, do not require information on the derivatives. This means that no information is lost, which additionally protects the algorithms against spurious solutions. Also, GA and SA offer a forward approach to solve the inverse problem. This fact makes them easily applicable to a range of optimization problems. Finally, although these algorithms are not guaranteed to find the global minimum, they provide a reasonable and stable approximation to the global optimum.

From the perspective of optimization schemes, the performance of numerical techniques applied is related to the degree to which the observed BOLD signal is approximated by the modeled response. A separate question of interest pertains to a performance measure that allows one to judge the quality of the reconstructed neural function associated with the modeled BOLD signal. Strictly speaking, there

is no guarantee that the reconstructed input is the best possible solution. However, the nature of global optimization techniques allows us to believe that the solution obtained through optimization is, to some degree, close to the true solution. In the case of simulations, we can directly evaluate discrepancy, or a distance, between the reconstructed and specified neuronal functions as the latter is known by construction of the testing scheme (see Fig. 2.2). What we need is just to define metrics (distances) in the appropriate function space. Within the framework of LFP simulations, the distance between two continuous functions, original and reconstructed, can be defined in a similar way as the discrepancy between the modeled and observed BOLD signals (see 2.9). Metrics can also be defined in a space of event sequences, allowing us to quantify similarity between spike trains in the case of the MUA simulations. Some metrics can correspond to Euclidian distances when spike trains are convolved with a smoothing kernel, by using the scalar product of the resulting functions (van Rossum, 2001). Cost-based metrics are another approach, wherein a cost is associated to insertions, deletions and shifts of spikes. The comparison of genetic sequences in Sellers (1974) and of electroencephalography (EEG) analysis in Wu and Gotman (1998) are examples of such an approach.

When the inverse methods are applied for analyzing real data, the true solution is unknown, and we can only indirectly judge the quality of the reconstructed neuronal activity underlying the BOLD signal. The stimulus timing waveform can only give us a general idea of what the true neuronal function looks like, but cannot claim to be a gold standard, especially if there are reasons to believe that the dynamics of neural activity should not be time-locked to the stimulus. Nevertheless, higher correlations

(smaller distances) between the reconstructed neuronal functions, smoothed or spiky, and the paradigm function may buttress the validity of the former. This issue is complicated by the necessity of defining the distance (metric) between continuous stimulus functions and discrete event sequences. A possible approach, as discussed above regarding MUA simulations, is to convolve spike trains with a smoothing kernel.

We do not claim that our methods are superior to previously proposed ones in the sense that our solution might be correlated with the stimulus waveform function to a higher degree. However, contrary to other proposed methods, we see an advantage in the possibility of finding the absolutely best input, provided two things are given: an observed BOLD response and a biophysical model taken for granted. This is possible by applying the techniques that are intrinsically designed to look for the global optimum. This implicitly assumes a secondary role for the stimulus function, with accent on the solution driven by the given model. A separate issue is the quality of the modeled physiological link between neural activity and BOLD responses, which is currently far from being mature. A better match of the reconstructed neural functions to the stimulus waveforms might be obtained with a better physiological model. Moreover, the idea of minimizing this discrepancy can guide future work and investigation of BOLD physiology.

As demonstrated by using simulations designed to test the forward-inverse relations between the neuronal dynamics and the evoked BOLD responses, the proposed methods allow one to correctly reconstruct an artificial underlying neuronal activity function. Furthermore, in an experiment with event-related design the numerical

techniques were able to capture the contour of a quite complicated paradigm waveform (see Fig 2.7). In addition, estimating some of the biophysical parameters of the model is possible simultaneously with reconstructing the neuronal input dynamics. However, it is quite challenging to put all the parameters into the optimization scheme. This is because some parameters are, by nature, scaling factors and some of these parameters can be balanced off by others. In fact, it is a multicollinearity problem, and some additional constraints need to be put on the possible values of the system parameters. Otherwise, clusters of possible solutions are expected. In general, for the purpose of inferring the underlying neural activity function in experiments with complicated design, it may be reasonable to run, as a preliminary stage, an experiment with a quite simple stimulus paradigm to estimate the biophysical parameters first. Next, the dynamics of the neural activity of interest can be reconstructed using the estimated parameters of the biophysical model.

An important issue in estimating the neural function and biophysical parameters is the question of whether we are dealing with a well-posed or ill-posed problem while trying to numerically reconstruct the input. In mathematics, all problems can be divided, roughly speaking, into two categories: well-posed and ill-posed problems. Let us consider the operator equation: $A\,u = h$,$u \in \mathcal{U}$, $h \in \mathcal{H}$, where $\mathcal{U}$ and $\mathcal{H}$ are metric spaces. According to Jacques Hadamard (Hadamard, 1902), the problem $A\,u = h$, $u \in \mathcal{U}$ is well-posed if (i) a solution u exists for any h; (ii) the solution is unique and (iii) and the solution u is stable with respect to small changes in h, *i.e.* the operator A^{-1} is defined for all $u \in \mathcal{U}$ and is continuous. Many optimal control problems, minimization problems, problems from linear algebra belong to

the class of ill-posed problem. In our case, the presence of noise in the BOLD signal can be considered as small perturbations of the function $h \in \mathcal{H}$. Riera et al. (2004) claim that the problem of reconstructing the underlying neural activity function is ill posed although they do not provide theoretical and numerical arguments in support of this claim. A possible approach to solve a ill-posed problem is to look for an approximate solution using regularization. From this perspective, the parametrization of the function $u(t)$ with B-splines (see (2.21)) can be interpreted as an implicit regularization for $u(t)$ or, in other words, an approximate criterion of the temporal variability of the neural function $u(t)$. We note that the possible ill-posedness of the optimization problem appears to be not a crucial issue when we use simulated annealing as an algorithm designed to find an approximate, although close to global, solution in the non-linear space. We were able to obtain approximately the same solution irrespective of initial guess (within a reasonable range) about the values of unknown parameters. Furthermore, the simulation tests with the various noise level in the BOLD signal $\hat{h}(t)$ reveal that bumps in the function $\hat{u}(t)$ (see Fig 2.6(a)) can be robustly discerned. At some point, a low signal-to-noise ratio of BOLD signals will cause the algorithms to fail to find a correct solution. Nevertheless, the algorithms tested on real BOLD signals, which contain, in general, serially correlated noise with long-range correlations rather than white noise, were able to robustly obtain stable and reasonable solutions.

The possibility of inferring the amplitude and timing of the underlying neural activity from BOLD dynamics may be beneficial in a number of applications in brain studies. First, it could be of particular interest in dynamic causal modeling (DCM)

(Friston et al., 2003), wherein the concept of effective connectivity (Friston, 1994) is combined with biophysics of the local responses. In particular, this approach may be of special use for those brain areas which are not directly locked to the stimulus, and therefore cannot be meaningfully associated with the stimulus waveform. DCM is an attempt to determine how effective connectivity patterns within networks of brain regions evolve over time under the influence of experimental manipulations. Methodologically, DCM is based on the idea to transfer estimation and inference about the coupling among neuronal populations in different brain regions from the level of observed data down to the neural level through a supplementary model. So it is a combination of two things. First, we have to transform the observed signals into neuronal activity. Second, we need a model of dynamic coupling between neuronal populations. The first part of the DCM approach could be a direct application of the techniques developed here to solve the hemodynamic inverse problem.

Theoretically, the methods proposed in this work allow one to reconstruct time courses of neuronal activity for the whole brain in analogy to reconstructing BOLD activity for every voxel. Moreover, the techniques developed to compute activation maps can be based on the inferred neuronal data, putting aside the issue of validity of the model. However, in practice it is not easy to estimate the dynamics of the neuronal activity for all the voxels from a data set due to high computational costs associated with solving the inverse hemodynamic problem. Nevertheless, it is possible to invert the BOLD responses for particular brain regions of interest.

In addition, the techniques considered for studying the forward-inverse relations can be applied in integrating schemes for multimodal neuroimaging such as fMRI

and electroencephalography (EEG) (Lemieux et al., 2001; Riera et al., 2004; Wan et al., 2006), which may help to overcome the limitations imposed by individual techniques in their ability to localize neuronal dynamics in time and space. For example, in an attempt to fuse EEG and fMRI data (Riera et al., 2004), interrelations between synaptic activity and hemodynamics were examined within the framework of an quasi-linear autoregressive model formulated to couple fast and slow subsystems based on the causal relations of the biophysical model (2.3)-(2.8). Considered separately, the EEG modality suffers from the problem of robustly estimating spatial distribution the sources of the electromagnetic fields. This inverse problem is known not to have a unique solution (Riera et al., 1998). Functional MRI provides relatively good localization of neuronal activity. We can hypothesize that since spatial maps of the dynamics of neuronal activity computed through solving the hemodynamic inverse problem puts additional constraints on the possible solutions of the distribution of the sources of electric currents, it might help in solving the electrophysiological inverse problem.

In conclusion, we have demonstrated that the neuronal activity underlying BOLD responses can be robustly inferred through non-linear optimization schemes. Specifically, it is possible to reconstruct both the neuronal activity function and the parameters of the biophysical model that determines the transitional dynamics of the physiological variables. In addition, the proposed methods allow one to characterize the neuronal signals from the perspective of what type of neuronal activity is best correlated with the hemodynamic responses: MUA or LFP. We expect that the proposed approach can refine the quantitative assessment and models of the physiological pro-

cesses underlying BOLD. In prospect, the techniques for solving the hemodynamic inverse model can be applied in dynamic causal modeling approach, allowing more meaningful inferences about effective connectivity. Finally, the proposed methods may potentially be applied for multimodal integration of neuroimaging techniques.

Bibliography

J.L. Boxerman, P.A. Bandettini, K.K. Kwong, J.R. Baker, T.L. Davis, B.R. Rosen, and R.M. Weisskoff. The intravascular contribution to fMRI signal change: Monte Carlo modeling and diffusion-weighted studies in vivo. *Magn Reson Med*, 34(1): 4–10, 1995a.

J.L. Boxerman, L.M. Hamberg, B.R. Rosen, and R.M. Weisskoff. MR contrast due to intravascular magnetic susceptibility perturbations. *Magn Reson Med*, 34(4): 555–66, 1995b.

R.L. Buckner. The hemodynamic inverse problem: making inferences about neural activity from measured MRI signals. *Proc Natl Acad Sci U S A*, 100(5):2177–9, 2003.

R.B. Buxton, E.C Wong, and L.R. Frank. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn Reson Med*, 39(6): 855–864, 1998.

J.M. Carnicer and J.M. Pena. Totally positive bases for shape preserving curve design and optimality of B-splines. *Comput. Aided Geom. Design*, 11(6):633 – 654, 1994.

R.W. Cox. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*, 29(3):162–73, 1996.

C. de Boor. *A practical guide to Splines*. Springer-Verlag, 1978.

K.J. Friston. Bayesian estimation of dynamical systems: an application to fMRI. *Neuroimage*, 16(2):513–30, 2002.

K.J Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Hum Brain Mapp*, 2:56–78, 1994.

K.J. Friston, P. Jezzard, and R. Turner. Analysis of functional MRI time-series. *Hum. Brain Mapp.*, 39:153–71, 1994.

K.J. Friston, O. Josephs, G. Rees, and R. Turner. Nonlinear event-related responses in fMRI. *Magn Reson Med*, 39(1):41–52, 1998.

K.J. Friston, A. Mechelli, R. Turner, and C.J. Price. Nonlinear responses in fMRI: the balloon model, volterra kernels, and other hemodynamics. *Neuroimage*, 12(4): 466–477, October 2000.

K.J. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–302, 2003.

F. Glover and G.A. Kochenberger. *Handbook of Metaheuristics*. Kluwer Academic Publisher, 2003.

R.L. Grubb, M.E. Phelps, and J.O. Eichling. The effect of vascular changes in PaCO2 on cerebral blood number, blood flow and vascular mean transit time. *Stroke*, 5: 630–9, 1974.

J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Bull. Univ. Princeton*, 13:49–52, 1902.

D.A. Handwerker, J.M. Ollinger, and M. D'Esposito. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage*, 21(4):1639–51, 2004.

D.J. Heeger and D. Ress. What does fMRI tell us about neuronal activity? *Nat Rev Neurosci*, 3(2):142–51, 2002.

D.J. Heeger, A.C. Huk, W.S. Geisler, and D.J. Albrecht. Spikes versus BOLD: what does the neuroimaging tell us about neural activity. *Nat. Neurosci.*, 3(7):631–3, 2000.

F.G. Hoogenraad, P.J. Pouwels, M.B. Hofman, J.R. Reichenbach, M. Sprenger, and E.M. Haacke. Quantitative differentiation between BOLD models in fMRI. *Magn Reson Med*, 45(2):233–46, 2001.

H.B. Keller. *Numerical methods for two-point boundary-value problems*. Blaisdell, 1968.

D.S. Kim, I. Ronen, C. Olman, S.G. Kim, K. Ugurbil, and L.J. Toth. Spatial relationship between neuronal activity and BOLD functional MRI. *Neuroimage*, 21(3):876–85, 2004.

L. Lemieux, A. Salek-Haddadi, O. Josephs, P. Allen, N. Toms, C. Scott, K. Krakow, R. Turner, and D.R. Fish. Event-related fMRI with simultaneous and continuous EEG: description of the method and initial case report. *Neuroimage*, 14(3):780–7, 2001.

M. Locatelli. Simulated annealing algorithms for continuous global optimization: Convergence conditions. *Journal of Optimization Theory and Applications*, 104 (1):121–33, 2000.

N.K. Logothesis, H. Merkle, M. Augath, T. Trinath, and K. Ugurbil. Neurophisiological investigation of the basis of the fMRI signal. *Nature*, 6843(412):150–157, 2001.

N.K. Logothetis. The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philos Trans R Soc Lond B Biol Sci*, 1424 (357):1003–37, 2002.

N.K. Logothetis and B.A. Wandell. Interpreting the BOLD signal. *Annu Rev Physiol*, 66:735–69, 2004.

J.B. Mandeville, J.J.A. Marota, C. Ayata, G. Zaharchuk, M.A. Moskowitz, B.R. Rosen, and R.M. Weisskoff. Evidence of cerebrovascular postarteriol windkessel with delayed compliance. *J.Cereb.Blood.Flow Metab.*, 19(6):679–89, 1999.

Z. Michalewicz. *Genetic algorithms + data structures = evolution programs*. Springer, 1996.

R. Mukamel, H. Gelbard, A. Arieli, U. Hasson, I. Fried, and R. Malach. Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science*, 309(5736):951–4, 2005.

D.G. Nair. About being BOLD. *Brain Res Brain Res Rev*, 50(2):229–43, 2005.

S Ogawa, T.M. Lee, A.R. Kay, and D.W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S A*, 87(24): 9868–72, 1990.

S. Ogawa, R.S. Menon, D.W. Tank, S.G. Kim, H. Merkle, J.M. Ellerman, and K. Ugurbil. Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging: a comparison of signal characteristics with a biophysical model. *Biophys.J.*, 3(64):803–812, 1993.

R. Parri and V. Crunelli. An astrocyte bridge from synapse to blood flow. *Nat Neurosci*, 6(1):5–6, 2003.

L.S. Pontryagin. *Mathematical Theory of Optimal Processes*. John Wiley & Sons, 1962.

S.B. Potter, P.J. Hakala, and M. Cropper. Stokes imaging of the accretion region in magnetic cataclysmic variables -I.Conception and realization. *Mon Not Astron Soc*, 297(4):1261–68, 1998.

J.O Ramsay and B.W. Silverman. *Applied Functional Data Analysis*. Springer, 2002.

G. Rees, K. Friston, and C. Koch. A direct quantative relationship between the functional properties of human and macaque V5. *Nat. Neurosci.*, 3(7):719–23, 2000.

C.R. Reeves and J.E. Rowe. *Genetic algorithms: Principles and Perspectives*. Kluwer, Norwell,MA, 2001.

J.J. Riera, M.E. Fuentes, P.A. Valdes, and Y. Oharriz. EEG-distributed inverse solutions for a spherical head model. *Inverse Problems*, 14:1009–19, 1998.

J.J. Riera, J. Watanabe, I. Kazuki, M. Naoki, E. Aubert, Ozaki.T., and R. Kawashima. A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signals. *Neuroimage*, 21(2):547–67, 2004.

G.E Sarty. *Computing brain activity maps from fMRI time-series images*. Cambridge University Press, 2007.

P.H. Sellers. On the theory and computation of evolutionary distances. *SIAM J Appl Math*, 26(4):787–93, 1974.

A. Shmuel, M. Augath, A. Oeltermann, and N.K. Logothetis. Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. *Nat Neurosci*, 9(4):569–77, 2006.

M. C. W. van Rossum. A novel spike distance. *Neural Comput*, 13(4):751–63, 2001.

A.L. Vazquez and D.C. Noll. Nonlinear aspects of the BOLD response in functional MRI. *Neuroimage*, 7(2):108–18, 1998.

X. Wan, J. Riera, K. Iwata, M. Takahashi, T. Wakabayashi, and R. Kawashima. The neural basis of the hemodynamic response nonlinearity in human primary visual cortex: Implications for neurovascular coupling mechanism. *Neuroimage*, 32(2): 616–25, 2006.

L. Wu and J. Gotman. Segmentation and classification of EEG during epileptic seizures. *Electroencephalogr Clin Neurophysiol*, 106(4):344–56, 1998.

R.L. Yang. Convergence of the simulated annealing algorithm for continuous global optimization. *Journal of Optimization Theory and Applications*, 104(3):691–716, 2000.

Y. Zheng, J. Martindale, D. Johnston, M. Jones, J. Berwick, and Mayhew J. A model of the hemodynamic response and oxygen delivery to brain. *Neuroimage*, 16(3):617–37, July 2002.

Chapter 3

Extracting chronometric information from fMRI signals through solving the hemodynamic inverse problem

3.1 Preliminaries

This chapter is based on the manuscript titled "Extracting information on mental chronometry from fMRI signals through solving the hemodynamic inverse problem" by Vakorin,V.A., Borowsky, R. and Sarty,G.E., submitted to *Neuroimage*. This work continues to probe issues raised in Chapter 2. Specifically, as shown in Chapter 2, mathematical models describing the transduction of neural activity into the BOLD response provide a framework for solving the hemodynamic inverse problem, *i.e.* reconstructing the dynamics of neural activity underlying the BOLD signal. Solving the hemodynamic inverse problem may potentially recover information that was hidden in the chain of events leading from the microscopic level of dynamic neuronal activity up to changes in hemodynamics.

The proposed techniques for recovering neural timing information (mental chronometry) are illustrated by using a neural model of an object interaction decision task.

The model assesses the involvement of semantic activation in the process of transferring information from the visual word/object system to the speech activation system. The dynamics of neural activity, associated with visual word/object processing, interaction semantics and speech were estimated based on the expanded "balloon" model, which mathematically describes the path from neuronal activity to the BOLD response through the changes in a vasoactive signal, cerebral blood flow, blood volume and deoxyhemoglobin concentration. Two aspects of the neural models were analyzed at both the hemodynamic and neural levels: effective connectivity and mental chronometry. We found that the causal path pattern was approximately the same at both levels. With mental chronometry, however, the neural function-based analysis has revealed that the main delay between visual word recognition and spoken word production occurs on the pathway leading from visual word activation to semantics. At the same time, the results of hemodynamic-based chronometric analysis were not able to reveal any significant latency. The comparative analysis has demonstrated the usefulness of mathematical modeling in the issue of increasing the temporal resolution of fMRI studies.

3.2 Introduction

Functional magnetic resonance imaging (fMRI) has emerged as a powerful tool to explore brain function. An important issue regarding the dynamics of brain activity is related to the question of how brain areas work together to carry out a specific task. Addressing this question has been based on two, not mutually exclusive but

rather complementary, tenets: functional specialization and functional integration (Sejnowski et al., 1988; Churchland and Sejnowski, 1988). The principle of functional specialization implies that distinct brain areas are specialized in their own functional roles, while the principle of functional integration rests on the assumption that different tasks are associated with different functional networks. The network paradigm of cortical representation encompasses the problem of specifying activated brain regions, representing processing elements interconnected with each other, and the strength of the connections between them. Two classes of neural modeling can be distinguished in the literature of neuroimaging studies: systems-level neural modeling and large-scale neural modeling (Horwitz et al., 1999, 2000; Horwitz, 2004).

Within the framework of systems-level modeling (McIntosh and Gonzalez-Lima, 1994a,b), the entire analysis is carried out at one level, *e.g.* at the hemodynamic level as in the case of fMRI data. Neuroanatomical information is exploited to define the functional networks, each unit of which is associated with the dynamics of the hemodynamic responses. One way to characterize such a system is to determine functional connectivity, based on correlating fMRI time-series related to spatially remote brain regions (Friston, 1994; Horwitz, 2003). Another way is to explore causal relationships between the neural units, and the influence some units exert over others. The latter approach incorporates the idea of effective connectivity (Friston, 1994), the concept of evaluating the task/context dependent pattern of the causal relationships between the neural units through direct and indirect effects, when a task is mediated by neural circuitry. Statistically, the issue of estimating the causal connectivity can be formulated as the problem of explaining covariances in interre-

gional activity through structural equation modeling (SEM), a statistical technique used for building and testing causal models (Buchel and Friston, 1997, 2000). Computationally, SEM is based on minimizing the discrepancy between the modeled and observed covariances.

Nevertheless, tracing the topographical structure of activation across brain regions is only one side of the story. SEM discounts temporal information; in particular, it discounts a possible time delay attributed to information transmission from one functional unit to another. Mental chronometry [1], originally defined in terms of reaction times (Posner, 1978), is an attempt to shed more light on neural dynamics by decomposing a cognitive or perceptual task into a temporal sequence of stages associated with the processing of information in different regions. Encouraging further investigations, the fMRI modality has recently proven to be informative in the issue of tracking the mental processing stages (Menon et al., 1998; Miezin et al., 2000; Formisano and Goebel, 2003). FMRI-based mental chronometry works on the assumption that the relative latency of neuronal events should somehow be expressed as a temporal shift between fMRI time series. A crucial concern about the suitability of functional MRI for the temporal analysis of the underlying neural activity is the limited temporal resolution of fMRI (Menon and Kim, 1999). Even with the ability to reach one hundred millisecond imaging resolution (Menon et al., 1998; Hernandez et al., 2002), the typical temporal resolution of fMRI studies remains on the order of a few seconds.

[1]Neural chronometry probably would be a better term since our goal is to recover latency in activation at the neural level. Nevertheless, we will follow the terminology of the paper by Menon et al. (1998)

In spite of some progress in analyzing fMRI data, that analysis (especially if it is of connectivity or chronometry) is compromised by one significant shortcoming: fMRI does not measure neuronal activity directly. Blood oxygen level dependent (BOLD) contrast fMRI signals, the most commonly used functional neuroimaging contrast (Ogawa et al., 1990), is not an exception. This contrast senses microscopic inhomogeneities in the magnetic field, caused by the presence of paramagnetic deoxygenated hemoglobin in red blood cells. BOLD response is a consequence of a chain of effects, depending on baseline and changes in blood volume, baseline and changes in blood flow, and changes in oxygenation (for reviews see Heeger et al. (2000); Heeger and Ress (2002); Logothesis et al. (2001); Logothetis (2002); Nair (2005); Sarty (2007)). Thus, blood flow regulation is the basis of the intrinsic physiologic limits of BOLD fMRI. The issue is complicated by the unclear nature of the neurovascular coupling (for review see Koehler et al. (2006); Haydon and Carmignoto (2006)). For example, a number of studies have emphasized the signaling role of astrocytes in controlling vasodilation, but not without controversy (Zonta et al., 2003; Takano et al., 2006).

To overcome the difficulties related to making inferences at the hemodynamic level, a number of studies applied large-scale neural modeling, an approach for modeling neurobiologically realistic interactions between brain regions, relating changes in hemodynamic activity to neural dynamics (Horwitz et al., 1999; Horwitz and Tagamets, 1999; Horwitz, 2004). This approach assumes that the regions are constructed of neural units, making a complex functional model, and that the neural units interact with each other through feedback and feed-forward connections. In addition, the large-scale models are multilevel, unifying modalities of variegated form, *e.g.*

extracellular field potentials and BOLD responses (Wan et al., 2006; Husain et al., 2004; Tagamets and Horwitz, 1998). Mathematical models, which can describe the physiological link between neural activity, associated with processing information, and corresponding changes in MR signals, may be considered as an intrinsic and essential part of the large scale models.

In spite of the absence of exact description of transduction of neural activity into BOLD signals, combined neuroimaging studies and electrophysiological recordings have demonstrated a coupling between BOLD and different types of neuronal activity (Logothesis et al., 2001; Heeger et al., 2000; Rees et al., 2000; Mukamel et al., 2005). The studies by Heeger et al. (2000) and Rees et al. (2000) have showed a positive correlation between behavioral measures, action potential and fMRI signals through indirect comparison between responses from single neuron recordings in monkeys and fMRI measurements in humans. More specifically, fMRI response in human cortical areas V1 and V5 was found to be proportional to the firing rate of individual neurons in the same areas in monkeys. Logothesis et al. (2001) have evaluated a model of linear coupling between simultaneously recorded fMRI signals and neuronal potentials in a study with the anaesthetized monkey with the electrode placed in the extracellular space in V1. Local field potentials (LFP), the low-frequency component of the extracellular field potentials, accounted for approximately 8% more in the variance of fMRI signals than did multi-unit spiking activity, a high frequency component. Despite the slightly better predictive power of LFPs, the possibility that spikes are a key determinant of the BOLD response, was not excluded (Logothetis, 2002). Further, through simultaneous fMRI and electrophysiological recordings in monkeys,

Shmuel et al. (2006) have observed that negative BOLD signal is associated with decreases in local-field potentials and multi-unit activity. The linear coupling between fMRI and neuronal activity in the auditory cortex of conscious humans was tested in the study by Mukamel et al. (2005). The authors found a high linear correlation between spiking activity, high-frequency LFPs and BOLD signals.

Although avoiding the issue of what types of neural activity are best correlated to hemodynamic responses, there have been a few attempts to mathematically model the chain of events leading from the voxel-aggregated neural activity to BOLD signals (Friston et al., 2000; Zheng et al., 2002). From the composition point of view, this link is represented as an input-output model, with the neural activity function acting as the input, and with the BOLD signal as the output. Mathematically, these models are described as state-space models consisting of process equations and an observation equation. The observation (measurement) equation defines the BOLD responses as a nonlinear function of physiological variables. The process equations determine the transitional dynamics of the physiological variables, including a vasodilation signal, cerebral blood flow, cerebral blood volume and deoxyhemoglobin content (Friston et al., 2000), and, in addition to these variables, in a more complicated model by Zheng et al. (2002), oxygen extraction fraction, mean capillary O_2 concentration and the ratio of average oxygen concentration in tissue over plasma oxygen concentration at the arterial end of the capillary.

Assuming that neuronal activity determines the dynamics of the physiological variables, and thus the shape of the BOLD response, we may infer the dynamics of the underlying neural activity. Friston (2002) has used quasi-analytical methods

based on Volterra series to reconstruct the neural activity function in terms of scaling coefficients for input paradigm waveforms through the expectation-maximization (EM) algorithm. Riera et al. (2004) have conducted the analysis within the framework of local linearization (LL) scheme. In the LL approach, the neural activity function was parameterized in terms of radial basis functions, so the unknown parameters were coefficients of their linear combination and, possibly, some of the biophysical parameters of the model. Computationally, the LL method belongs to a class of recursive filters reduced to linearized equations of evolution for the conditional means of the state variables and their covariance matrix. Both the EM and LL approaches do not address the issue of whether they can reproduce the global solution to the hemodynamic inverse problem, avoiding spurious local optima.

Vakorin et al. (2006) have proposed to transform the original forward state-space model by Friston et al. (2000) into an optimal control problem with the objective of minimizing the discrepancy between observed and modeled BOLD signals over all admissible neural activity functions. Adopting this framework turned out to be fruitful in the issue of interpreting the meaning of the neuronal function used in the model. Specifically, applying the Pontryagin minimum principle (Pontryagin, 1962), it was shown that off/on dynamics of neural activity is the natural mathematical solution of the model. In particular, a piece wise step paradigm function, traditionally representing block design or event-related design, is a specific case of the theoretical solution. Alternatively, the on/off neural dynamics can be interpreted as a spike train, which can be associated with multi-unit spiking activity. In addition, global optimization techniques, such as genetic algorithms and simulated annealing, can be

applied to find the best control input: the neuronal activity and/or the biophysical parameters used in the model.

Solving the hemodynamic inverse problem, *i.e.* reconstructing the timing and amplitude of the neural signal underlying the BOLD responses, provided that the intrinsic path is specified, can potentially produce new information. One way to produce new information is to use a multimodal approach, combining electrophysiological recording and fMRI data, which may allow us to refine the nature of physiological link between the neural and hemodynamic levels when the data are compared to predictions from neural-hemodynamic mathematical model. Another way to extract new information is the possibility of recovering, not exploiting data from another modality, extra information from fMRI time series per se, hidden when one goes from the microscopic level of neuronal activity to the sluggish changes in the hemodynamic signal.

In the present study, by an example of a functional network model from cognitive neuroscience, we will illustrate how one can infer information on mental chronometry through solving the hemodynamic inverse problem. Specifically, relative timing differences in information processing between regions of interest will be estimated by shifting the stress of calculations from original fMRI data to reconstructed neural dynamics. The functional model used in the present study is based on a dorsal-ventral model of visual information processing, whereby the identification of a word or object is processed primarily by the ventral stream (i.e., occipital-temporal cortex), whereas deciding how to interact with an object is processed primarily by a mid-dorsal stream (i.e., occipital-parietal-frontal cortex; see Milner and Goodale (1995);

Borowsky et al. (2005)). Using an object interaction decision task with word stimuli (from Borowsky et al. (2005)), we assume the involvement of visual word/object processing, interaction semantics and speech systems. From the neuroscience point of view, one might expect regional latencies that lead to different relative delays in processing information in different functional regions. Specifically, the models consider the visual word/object module as the starting point in a sequence of activation, *i.e.* a word stimulus will likely evoke an image of the object that it refers to given the task of deciding how one interacts with it. Furthermore, there are two possible pathways from visual word recognition to spoken word production: a direct pathway from the visual word/object system to the speech activation system (reflecting highly familiar noun-verb associations, such as "ball-throw"), and a less direct route, from the visual word/object system to the interaction semantics, and then, from semantic system to speech activation (reflecting less familiar associations between objects and possible interactions with them,*e.g.*, "jeans - wear them").

For the present study, we will assess the pattern of effective connectivity between functional areas associated with the visual word/object system, the interaction-semantics knowledge system and the speech system, testing the strength of causal influence through two information processing pathways from visual word recognition to the speech system, direct and indirect. Further, we will attempt to estimate the relative delays in information processing between the three functional modules. The rational for this is to trace both topography and the sequence of cortical activation. Next, based on the observed fMRI time series attributed to the functional modules, we will solve the hemodynamic inverse problem, and then, at the neural level, we

will repeat the testing procedures pertaining to effective connectivity and mental chronometry. Our aim is to explore the possibility of recovering more information from the dynamics of fMRI signals *per se*, associated with the neural units of the network, by going down to the neural level from the hemodynamic level through the intrinsic pathway defined by a biophysical mathematical model.

3.3 Methods

3.3.1 Hemodynamic inverse problem

The BOLD effect is a complex physiological phenomenon. Mathematical models (Buxton et al., 1998; Zheng et al., 2002) have been recently proposed to link evoked fMRI BOLD responses to underlying neural signals. The expanded balloon model (Buxton et al., 1998) describes the coupling between neural activity and the BOLD signal in terms of the dynamics of four physiological variables, represented by the vector $\boldsymbol{x}(t)^T = (x_1(t), x_2(t), x_3(t), x_4(t))$, namely

$$
\begin{aligned}
x_1(t) &= s(t) \ = \text{the flow-inducing signal} \\
x_2(t) &= f_{in}(t) \ = \text{regional cerebral blood flow (rCBF)} \\
x_3(t) &= v(t) \ = \text{regional cerebral blood volume(rCBV)} \\
x_4(t) &= q(t) \ = \text{deoxyhemoglobin concentration}
\end{aligned}
$$

All quantities are normalized relative to their resting values so that all the variables represented by the vector $\boldsymbol{x}(t)$ are dimensionless.

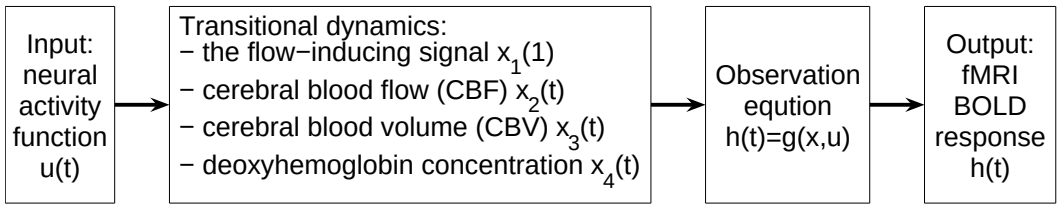


Figure 3.1: A diagram illustrating the structure of the expanded balloon model. It is an input-output hierarchal model: the neuronal activity function is the input at the lower level, and the BOLD response is the output at the higher level. The intrinsic biophysical system is formulated as a state-space model. The process equations describe the transitional dynamics of four physiological variables: a vasoactive signal, cerebral blood flow (CBF), cerebral blood volume (CBV) and deoxyhemoglobin content. In the measurement equation, the BOLD signal is partitioned into extravascular and intravascular components, defined in terms of deoxyhemoglobin concentration and CBV.

Biophysically, the expanded balloon model is based on a tandem connection of three successive parts: a model for the BOLD signal, based on the paramagnetic properties of deoxygenated hemoglobin, affecting local susceptibility; the balloon component, a model of capillary compartments, which can be mechanically expanded or constricted under the influence of cerebral blood flow; and the rCBF component, a dynamic model describing changes in the blood flow in reaction to neural activity mediated through the release of a vasoactive agent. Mathematically, the model is expressed as a space-state model with two sets of equations: measurement and process equations (see Fig. 3.1). The measurement (observation) equation reflects the dynamics of the BOLD signal $h(t)$ as a non-linear function of deoxyhemoglobin concentration and rCBF, modeling extravenous and intravenous components weighted by relative volumes:

$$h(t) = g\left(\boldsymbol{x}(t), u(t)\right) = V_0\left(k_1(1 - x_4(t)) + k_2\left(\frac{1 - x_4(t)}{x_3(t)}\right) + k_3(1 - x_3(t))\right) \quad (3.1)$$

The process equations portray the transitional dynamics of the flow-inducing (va-

soactive) signal, rCBF, rCBF and deoxyhemoglobin concentration through a system of non-autonomous (with the exogenous input $u(t)$), first order differential equations

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = f_1(\boldsymbol{x}(t), u(t)) = u(t) - \frac{x_1(t)}{\tau_s} - \frac{x_2(t) - 1}{\tau_f} \tag{3.2}$$

$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = f_2(\boldsymbol{x}(t), u(t)) = x_1(t) \tag{3.3}$$

$$\frac{\mathrm{d}x_3}{\mathrm{d}t} = f_3(\boldsymbol{x}(t), u(t)) = \frac{1}{\tau_0}\left(x_2(t) - x_3(t)^{\frac{1}{\alpha}}\right) \tag{3.4}$$

$$\frac{\mathrm{d}x_4}{\mathrm{d}t} = f_4(\boldsymbol{x}(t), u(t)) = \frac{1}{\tau_0}\left(\frac{x_2(t)}{E_0}\left[1 - (1 - E_0)^{\frac{1}{x_2(t)}}\right] - x_4(t)x_3(t)^{\frac{1-\alpha}{\alpha}}\right) \tag{3.5}$$

where $\Theta \equiv \{\tau_s, \tau_f, \tau_0, \alpha, E_0, V_0, k_1, k_2, k_3\}$ is the set of biophysical parameters, and the neural activity function $u(t)$ acts as an input into the model for the rCBF component. The initial condition for the set of differential equations (3.2)-(3.5) is

$$\boldsymbol{x}^\top(0) = (x_1(0), x_2(0), x_3(0), x_4(0)) = (0, 1, 1, 1) \tag{3.6}$$

The biophysical parameters of the model are the time constant of the flow-inducing signal decay τ_s, the time constant of the feedback regulatory mechanism τ_f, the mean transit time of a blood cell in the venous compartment τ_0, Grubb's exponent α for the flow-volume relation, baseline oxygen extraction rate E_0, and resting blood volume fraction V_0 (for more about these parameters see Friston et al. (2000)). The dimensionless parameters k_1, k_2 and k_3 depend on experimental and physiological parameters, and reflect the effects of rCBF and deoxyhemoglobin concentration on the BOLD signal. The parameters E_0, α, V_0 are dimensionless as well, contrary to the parameters τ_s, τ_f and τ_0 that are measured in units of time.

In a forward approach, the neuronal activity $u(t)$ determines the dynamics of the physiological variables $\boldsymbol{x}(t)$, and thus the shape of the hemodynamic response $h(t)$.

This relationship can be used to solve the hemodynamic inverse problem: the issue of reconstructing the neuronal activity underlying the BOLD signal. In an approach proposed by Vakorin et al. (2006), the neuronal function $u(t)$ is to be chosen in order to minimize discrepancy between the model and experiment, a distance $D(h, \hat{h})$ between two functions, namely between the observed BOLD $\hat{h}(t)$ and the function $h(t) = g\left(\boldsymbol{x}(t), u(t)\right)$ as predicted through the transitional dynamics (3.1)-(3.5) for a given $u(t)$. Specifically, $u(t)$ is to be found from

$$\hat{u}(t) = \arg\min_{u(t)} D(h, \hat{h}) \tag{3.7}$$

where the distance D between two functions is defined in L^2 sense, as

$$D(h, \hat{h}) \equiv D(u) \equiv \int_0^T \Big(\hat{h}(t) - g\left(\boldsymbol{x}(t), u(t)\right) \Big)^2 \mathrm{d}t \tag{3.8}$$

Further, the function $u(t)$ can be parameterized in the terms of a vector $\boldsymbol{\theta}$ of parameters, assumed to be unknown, and a set $\boldsymbol{\Phi}$ of known basis functions:

$$u = u(t, \boldsymbol{\theta}, \boldsymbol{\Phi}) \tag{3.9}$$

Then, the problem defined in (3.7) becomes an optimization problem over $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \int_0^T \Big(\hat{h}(t) - g\left(\boldsymbol{x}(t), \boldsymbol{\theta}\right) \Big)^2 \mathrm{d}t \tag{3.10}$$

subject to (3.1)-(3.6).

Further, the problem of reconstructing the neuronal activity underlying the BOLD signals from repeated block experiments can be fused with the idea of signal averaging, a technique known to increase signal-to-noise ratio of fMRI signals (DeYoe et al., 1994). Specifically, we assume that the fMRI time series consist of P repeated

imaging blocks (periods), with a time period of T_0, so the total time of the experiment is $T = T_0 \times P$. Let the function $u_0(t)$ represent the neuronal activity function averaged over P blocks. This function $u_0(t)$ can be repeated P times, which produces a function $u(t)$ spanning the entire experiment:

$$u = \mathcal{A}_P \, u_0 \tag{3.11}$$

where $\mathcal{A}_P$ is a P times repeat operator such that $\mathcal{A}_P : L_{[0,T_0]} \rightarrow L_{[0,P\times T_0]}$ ($L_{[a,b]}$ denotes the space of Lebesgue-integrable functions defined on the interval $[a, b]$) with $u(t + nT) = u_0(t)$ for $n \in \{0, \ldots, P - 1\}$.

In a number of studies, the local field potential (LFP), a low-frequency component of the extracellular field potential, is considered a better indicator of changes in BOLD signal, when compared to multi-unit spiking activity (Logothesis et al., 2001). Mathematically, a low-pass filter can be implemented through parameterizing the neural function $u_0(t)$ in a basis of known smooth functions,$\phi_j(t)$, as

$$u_0(t) = \sum_{j=1}^{K} c_j \, \phi_j \, (t) \tag{3.12}$$

where $\boldsymbol{c}^T = (c_1, \ldots, c_K)$ represents a vector of unknown coefficients. In addition, to model the delay between neuronal activity and the evoked hemodynamic response, the function $u_0(t)$ is allowed to move along the time axis, which can be represented by a shifting parameter d:

$$u_0(t) = \sum_{j=1}^{K} c_j \, \phi_j \, (t - d) \tag{3.13}$$

So, the vector $\boldsymbol{\theta}$ of unknown parameters consists of the coefficients $\boldsymbol{c}$ of the linear expansion (3.12) and the parameter d: $\boldsymbol{\theta}^T = \{\boldsymbol{c}^T, d\}$.

The problem defined in (3.1)-(3.6) and (3.10) is a multi-dimensional minimization problem over the space spanned by $\boldsymbol{\theta}$. The non-linearity of the optimization problem requires applying numerical techniques to estimate $\boldsymbol{\theta}$. Then, provided that the optimal parameters $\hat{\boldsymbol{\theta}}$ are estimated, the dynamics of the neuronal activity $\hat{u}(t) = u(t, \hat{\boldsymbol{\theta}}, \boldsymbol{\Phi})$ can be evaluated in the span of the shifted basis functions $\boldsymbol{\Phi}$.

3.3.2 Effective connectivity

Inferences regarding effective connectivity can be studied within the framework of structural equation modeling, a statistical technique for testing the hypothesis that some neural units exert causal influences over other units, through minimizing the discrepancy between the actual covariance matrix and that specified by the fitted model. Our model of reading a word and deciding how to interact with its referent object is composed of three functional modules, allowing for visual word/object processing (V), interaction semantics (I) and speech (S). The path coefficients β_{VI}, β_{VS} and β_{IS} represent an estimate of the causal influence. One path of the model is the direct effect of the visual word/object system on speech production (β_{VS}). The other causal path is partitioned into the effects of processing in the visual word system on interaction semantics (β_{VI}) and the subsequent effect of interaction semantics on speech (β_{IS}). The SEM can be represented as a circuit diagram with three nodes and three arrows to designate unidirectional connections (see Fig. 3.2)

Let the functions η_V, η_I and η_S represent signals (time series or functions), at the neural or hemodynamic levels, associated with visual word/object processing, interaction semantic knowledge system and speech, respectively. As the diagram in

Fig. 3.2 connotes, there are three equations (describing the relationships between three covariances) and three unknowns (three path coefficients). The model can be defined in terms of multivariate regression. It can be shown (Keith, 2006) that the causal path coefficients β_{VI}, β_{VS} and β_{IS} can be estimated as the standardized regression coefficients in the following model

$$\begin{pmatrix} \eta_I \\ \eta_S \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \beta_{IS} & 0 \end{pmatrix} \begin{pmatrix} \eta_I \\ \eta_S \end{pmatrix} + \begin{pmatrix} \beta_{VI} \\ \beta_{VS} \end{pmatrix} \begin{pmatrix} \eta_V \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \tag{3.14}$$

where ε_1 and ε_2 are vectors of zero mean, uncorrelated random errors with common variance $\sigma^2_{\varepsilon_1}$ and $\sigma^2_{\varepsilon_2}$, $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\varepsilon}\mathbf{I}_N)$.

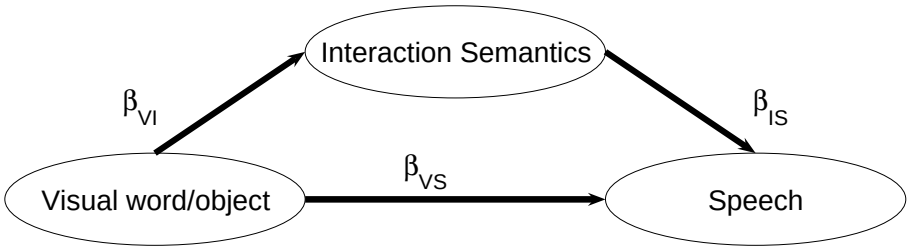


Figure 3.2: The structural equation model that represents a model of an object interaction decision task. The model is illustrated as a circuit diagram with three functional modules, allowing for a visual word/object system, an interaction semantic knowledge system and a speech production system. The arrows model the directions of causal influence between three modules. Specifically, the information can flow through two pathways: directly from visual word/object processing to speech along a ventral route, and, in a less direct way, from the visual word/object system to the interaction semantic knowledge system, and then, from the interaction semantics system to speech production along a mid-dorsal route. The coefficients β_{VI}, β_{VS} and β_{IS} reflect the strength of the causal influence one unit exerts over another.

The solved paths are an estimate of effective connectivity, describing the effects of the visual word/object system and the interaction-semantics system on speech, along with the effects of visual word/object processing on interaction semantics.

Since we are taking the parameters β_{ij} to the standardized path coefficients, they can be interpreted in standard deviation units. For example, a one percent increase in standard deviation of η_V will result in a β_{VI} standard deviation increase in η_I and a β_{VS} standard deviation increase in η_S. It should be remembered, however, that path analysis is a confirmatory technique. The validity of any effective connectivity found is limited by the validity of the predefined model.

3.3.3 Mental latency

One approach for assessing relative latency in processing information between two brain areas is to apply cross-correlation analysis. Regarding fMRI-driven studies, Bandettini et al. (1993) have computed the cross-correlation coefficients of the data with respect to a reference waveform to create activation maps. Saad et al. (2003) have used the Hilbert transform algorithm, based on an idea of cross-correlation, to estimate the delay of the BOLD response to neuronal stimulation.

Let the indices i and j denote regions of interests (ROI). Specifically, define the set Ω to include the brain areas associated with visual word processing (V), interaction semantics (I) and speech (S): $i, j \in \Omega$, where $\Omega \equiv \{V, I, S\}$. The cross-correlation coefficient $R_{ij}(\tau)$ between two real-valued functions, $\eta_i(t)$ and $\eta_j(t)$, is defined as

$$R_{ij}(\tau) = \int \eta_i(t)\eta_j(t+\tau)\,\mathrm{d}t \tag{3.15}$$

where the integral is over the appropriate values of the time variable t for a fixed time constant τ. In other words, the cross-correlation is defined in terms of one signal, shifted and multiplied by another signal. The time delay Δ_{ij} between two

signals can be estimated through the optimization problem

$$\Delta_{ij} = \max_{\tau} R_{ij}(\tau), \tag{3.16}$$

which defines the relative timing difference between two signals in terms of a shift between the signals that maximizes their cross-correlation function.

For the present study, we will compute first the cross-correlation functions between fMRI time series, $\hat{h}_i(t)$ and $\hat{h}_j(t)$, associated with regions $i, j \in \Omega$ of interest that are involved in an object interaction decision task. Further, for the same ROIs, we will calculate the cross-correlation coefficients between the reconstructed neuronal activity functions, $\hat{u}_i(t)$ and $\hat{u}_j(t)$ where $i, j \in \Omega$, which are defined as the solutions to the hemodynamic inverse problem (3.10) based on the original fMRI time courses for each $i \in \Omega$. Both at the neural and hemodynamic level, the cross-correlation functions were maximized to produce an estimate of signal delay.

3.3.4 Experiment

Six healthy subjects of age between 23 and 41 with a mean age of approximately 30 years, four female and two male, participated in this study. The study was approved by the University of Saskatchewan Behavioral Sciences Ethics Committee, and was conducted according to relevant institutional guidelines.

All participants performed a word object interaction decision task. Specifically, the stimuli were visually presented words. A word was displayed on a back-projection screen, and the participants were asked to overtly describe how they would interact with the word presented on the screen. For example, when they see the word “jeans”,

they could say something like "wear them". Image acquisition was synchronized with the visually presented stimulus and triggered by EPrime software (Psychology Software Tools, Inc., Pittsburgh, USA). Stimuli were projected with an LCD projector interfaced with the computer running the EPrime software. A mirror was attached to the MRI head coil to make the back-projection screen visible to the participant.

The imaging was performed with a 1.5 T Siemens Symphony (Erlanger, Germany) scanner. A gradient echo T_2^* single-shot EPI sequence (T_E = 55ms and T_R = 3300ms) was used with fat saturation pulses. A 1650 ms gap of no acquisition in each repetition ($\frac{1}{2}$ of T_R) was used. The 12 acquired axial slices had a square field of view of 250 mm and were 8-mm thick with an interslice gap of 2 mm. The 4^{th} or 5^{th} inferior-most slice was centered on the posterior commissure in order to image the entire cortex with one volume. The original 64×64 data matrices were Fourier reconstructed to 128×128 images. One presentation block contained 16 volumes, with the stimulus presented for the first 8 volumes, followed by 8 volumes of rest. The block pattern was repeated 5 times ($P = 5$). The first 6 image volumes, collected before the beginning of the first block to allow the spins to reach a steady state, were discarded prior to analysis. So the complete data set consisted of 86 image volumes of which 80 were used in the analysis. Time series with intensities below a cut-off of 200 grey-scale units (the black background) were also excluded from further analysis.

The brain regions associated with processing visual words/objects, semantics, and phonology/speech were identified based in neuroimaging reviews of object and language processing (Borowsky et al., 2005). In the review by Demb et al. (1999) on single word processing with fMRI and positron emission tomography (PET),

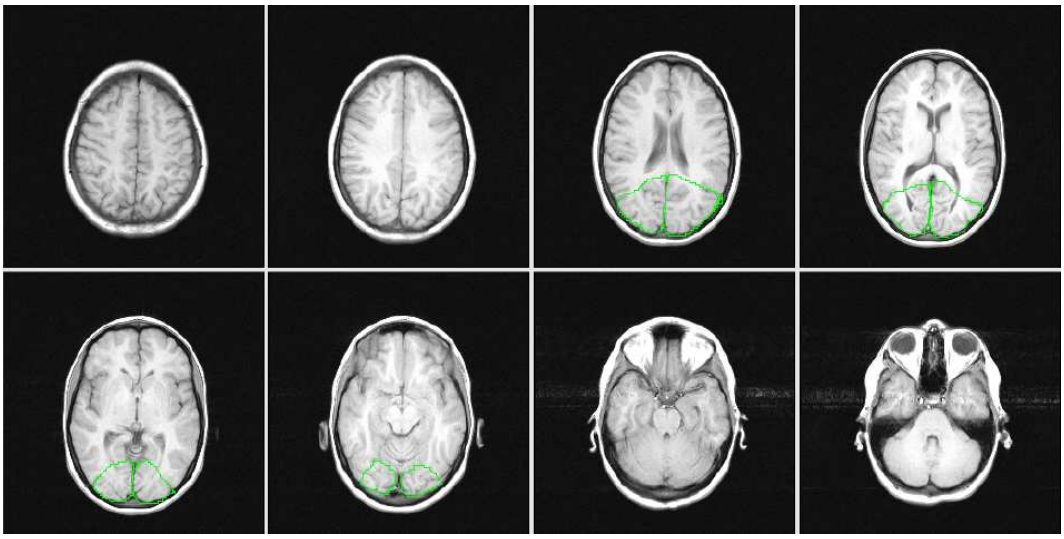


Figure 3.3: Masked here are the medial extratriate cortices associated consistently with processing visual word/object information (the region V).

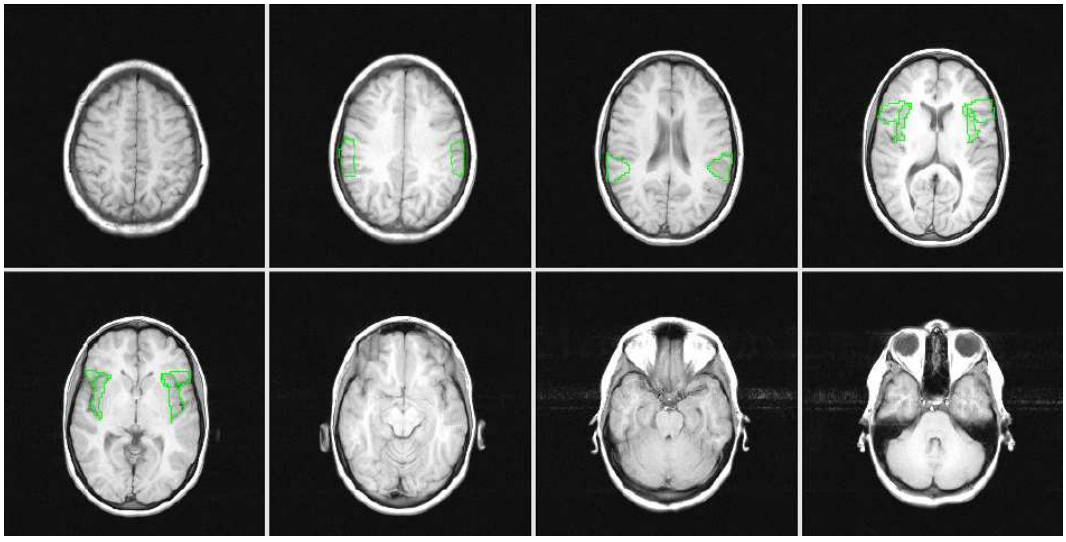


Figure 3.4: The inferior frontal gyri and insular cortices (the region S). These regions were reported in the literature to be related to speech processing of visually presented stimuli.

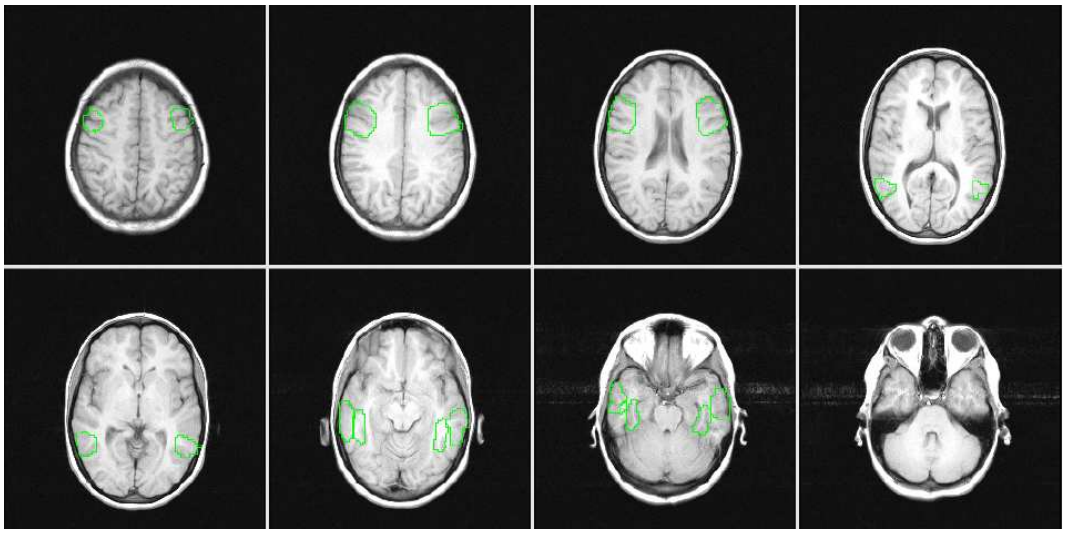


Figure 3.5: The lateral prefrontal cortex, middle temporal gyri and fusiform gyri (the region I) masked in an axial view. These regions have been identified in neuroimaging literature on language processing as regions associated with semantic processing of visually presented stimuli

the medial extratriate cortices tended to be related to visual word form. These regions consistently demonstrated activation to various control stimuli, including basic visual shape information (see also the review by Kanwisher et al. (2001)). Demb et al. (1999) have identified the pars opercularis and pars triangularis of the inferior frontal gyri as regions related to phonology/speech processing of visually presented words. Tasks requiring phonology/speech analysis relative to various control tasks have differentially revealed activation in these regions. In addition, Binder and Price (2001) and Borowsky et al. (2006) have emphasized insular activation associated with phonology/speech. Regarding the semantic processing of visually presented words, Demb et al. (1999) have reported activation in a lateral prefrontal region just superior to the inferior frontal gyrus. The same finding was supported in a review by Martin (2001) on semantic memory. In addition, the middle temporal gyri and

fusiform gyri have been identified as regions that are involved in the retrieval of semantic information (Martin, 2001; Palmer et al., 2004).

The following analysis was carried out for each participant separately. A correlation-based method, BOLDfold, was used to compute activation maps (Sarty and Borowsky, 2005). Mathematically, the BOLDfold approach is equivalent to an analysis of variance approach (Clare et al., 1999). In BOLDfold, the activation maps are calculated voxelwise based on the correlations between the intensity time course and its repeated mean time course. The correlation range above a correlation threshold of 0.65 was set to define the activated voxels. After the brain regions associated with visual word/object, interaction-semantics and speech/phonological processing had been masked, the average time series of all activated voxels within each ROI was generated. Thus, there was one averaged time series associated with activation of the visual word/object processing, interaction semantics and speech/phonology for each participant.

These averaged fMRI signals were used in two ways. First, they were analyzed to make inferences about mental chronometry and effective connectivity at the hemodynamic level. Second, these signals were used as the observed functions $\hat{h}(t)$ in the functional (3.8) to solve the inverse hemodynamic problem.

As a pre-processing step to solving the hemodynamic inverse problem, the averaged fMRI signals were normalized with respect to the mean intensity value averaged over the last four dummy volumes acquired to achieve the steady state before the main body of the experiment. In addition, any linear trend was eliminated from these signals. To alleviate possible edge effects related to integrating the system

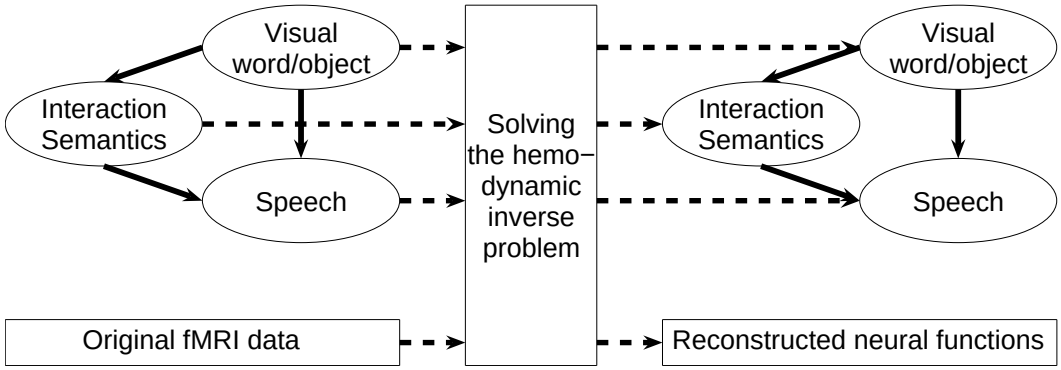


Figure 3.6: A scheme that illustrates the idea of going from the hemodynamic (fMRI) level down to the neural level through solving the hemodynamic inverse problem. Recently proposed mathematical models define the intrinsic path from neural activity to hemodynamic responses. As the neuronal activity determines the transitional dynamics of the physiological variables, and thus the shape of the BOLD response, we may infer the dynamics of the underlying neural activity. These methods are illustrated by using a cognitive neuroscience model of reading a word and deciding how to interact with its referent object. Specifically, the inverse problem was solved for each participant, based on the fMRI time series averaged with the regions associated with the visual word/object system, interaction semantic knowledge system and speech activation system.

(3.2)-(3.5) of differential equations, the time span of the experiment was extended by adding six zero values to the normalized BOLD time course before and two zero values after the main body. To increase the sampling rate of the BOLD signal for the purpose of numerically integrating the transitional system (3.2)-(3.5), the discrete adjacent points of the BOLD time series were linked together by linear splines.

The following values of the biophysical parameters were used in the simulations: $\tau_s = 0.8\,\mathrm{s}$, $\tau_f = 0.4\,\mathrm{s}$ and $\tau_0 = 1\,\mathrm{s}$ (Friston et al., 2000); $\alpha = 0.4$ (Grubb et al., 1974; Mandeville et al., 1999); $E_0 = 0.6$ which is within the range of the values of 0.4 and 0.8 as used in Buxton et al. (1998) and in Friston et al. (2000), respectively. As has been previously estimated for 1.5 T and T_E=40 ms, $k_1 \simeq 7E_0$, $k_2 \simeq 2$ and $k_3 \simeq 2E_0 - 0.2$ (Boxerman et al., 1995; Ogawa et al., 1993).

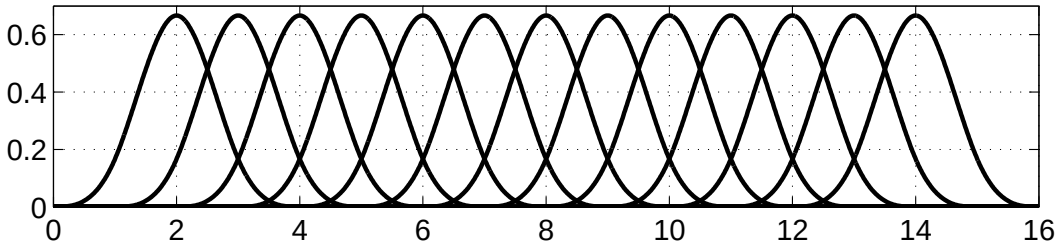


Figure 3.7: The set of B-spline functions used to put constraint on the smoothness of the reconstructed neural functions.

The hemodynamic inverse problem (3.10) was solved, based on the normalized averaged fMRI series associated with visual word/object identification, speech processing and semantic processing. Separately, for each ROI and for each participant, the discrepancy (3.10) between the expected and observed BOLD was minimized over the function $u_0(t)$, the averaged aggregated neural activity defined in (3.13). The function $u_0(t)$ was parameterized in terms of B-splines (de Boor, 1978), illustrated

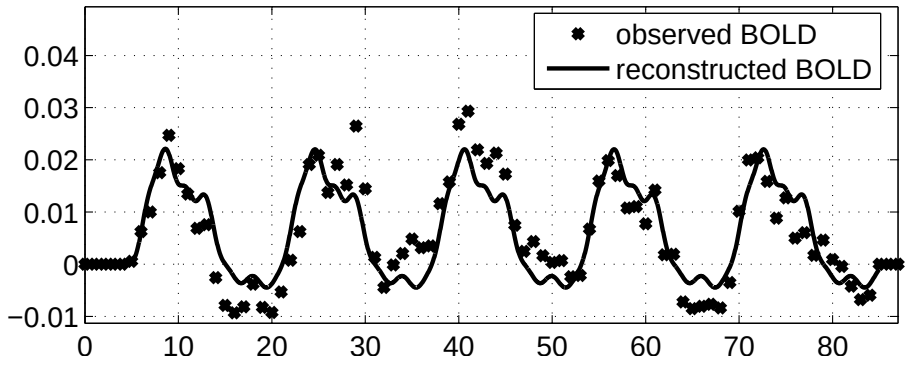
in Fig. 3.7. B-splines were chosen because of their stable numerical properties, their continuous derivatives at the joining points and their local support which prevents one observation from affecting the entire shape of the hemodynamic response. A 4^{th} order Runge-Kutta algorithm (Kalitkin, 1978) was applied to numerically integrate the differential system (3.2)-(3.5). Regarding optimization, we used simulated annealing (SA), for review see Glover and Kochenberger (2003). SA is a local search method capable of escaping from local minima by using limited hill-climbing steps, moves which increase the objective function value. It is based on the idea of taking a random walk at successively lower temperatures, where the probability of taking a step up is given by a Boltzmann distribution.

At this stage, for each subject there were three original fMRI times series and three reconstructed neural activity functions associated with visual word/object identification, speech processing and semantic processing. Causal path coefficients, β_{ij}, where $i, j \in \Omega$, were estimated through solving the multivariate regression defined in (3.14). Effective connectivity for the functional network defined in Fig. 3.6 was determined in terms of path coefficients averaged across six participants.

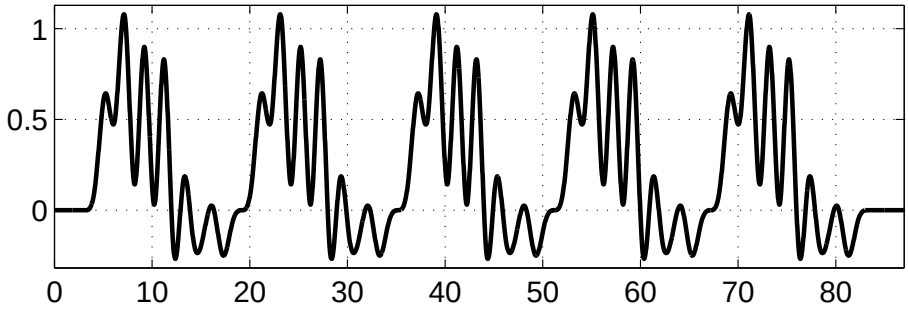
3.4 Results

Fig. 3.8 illustrates an example of solving the hemodynamic inverse problem based on the normalized original fMRI time course $\hat{h}(t)$ attributed to processing semantic information (for one subject). The observed discrete and noisy BOLD signal(Fig. 3.8(a), crosses) is superimposed on the modeled hemodynamic response

$h(t)$(Fig. 3.8(a), solid line) associated with the reconstructed neural activity $u(t)$ (Fig. 3.8(b)).



(a) BOLD responses, observed and modeled

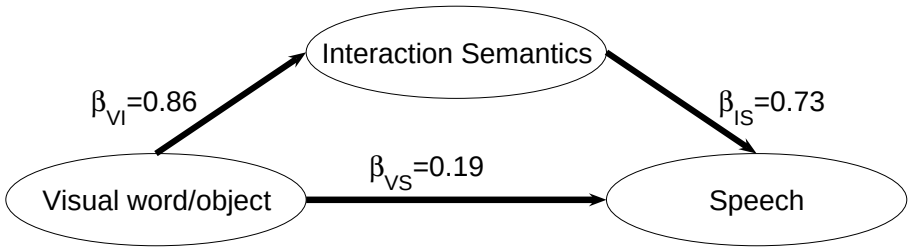


(b) Reconstructed neural activity function

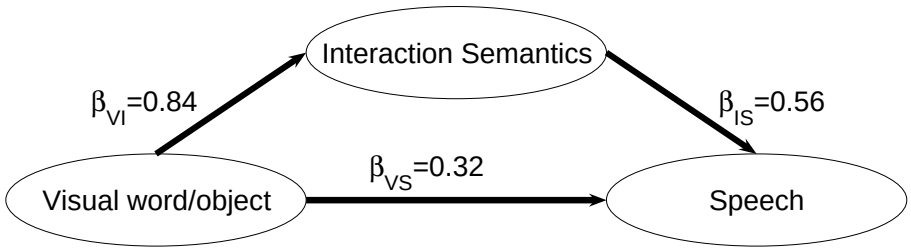
Figure 3.8: The dynamics of normalized and trend-eliminated BOLD signal and the reconstructed neural activity function averaged over 5 blocks, for the region associated with semantic processing for one of 6 participants: the expected, related to the reconstructed neural function (Fig. 3.8(b)), BOLD dynamics superimposed on the original time course (a); the estimated function $u(t)$ as a combination of the B-splines basis functions (b). The neural function $u(t)$ is constructed through unfolding the neural function that represents neural activity averaged over 5 periods.

Fig. 3.9(a) represents an estimate of causal influences at the hemodynamic level, when the variables η_V, η_I and η_S in (3.14) were represented by the original fMRI time series. We found that the average values are as follows: $\beta^h_{VI} = 0.86$, $\beta^h_{VS} = 0.19$ and $\beta^h_{IS} = 0.73$. All the coefficients were found significantly different from zero

with p-values less than 0.01. To determine effective connectivity at the neural level, the reconstructed neural functions $u(t)$ were used to estimate the path coefficients: $\beta^u_{VI} = 0.84$ (p-value less than 0.001), $\beta^u_{VS} = 0.32$ ($p = 0.14$) and $\beta^u_{IS} = 0.56$ ($p = 0.024$), which is illustrated in Fig. 3.9(b). The results obtained at the hemodynamic level supports the idea that information flows from visual word/object system to speech production via interaction-semantic knowledge system. At the neural level, the obtained results reveal approximately the same effective connectivity pattern.



(a) SEM at the hemodynamic level



(b) SEM at the neural level

Figure 3.9: Structural equation models (SEMs) of an object interaction decision task. The effective connectivity pattern was estimated in two ways: at the hemodynamic level, when the variables used to estimate the structural equation model in Fig. 3.6 were represented by the original fMRI time series (a); and at the neural level, when the variables associated with the visual word/object system, interaction semantic knowledge system and speech production were represented by the reconstructed neural activity underlying the observed hemodynamic responses (b). All the path coefficients, except for β_{VS} at the neural level, were found to be significantly different from zero ($\beta > 0$).

(a) Estimate of the onset latency at the hemodynamic level

path	mean ± std.error, T_R	t-statistics	p-value
Δ^h_{VI}	0.09 ± 0.05	1.83	0.13
Δ^h_{VS}	0.16 ± 0.14	1.15	0.30
Δ^h_{IS}	0.03 ± 0.05	0.59	0.58

(b) Estimate of the onset latency at the neural level

path	mean ± std.error, T_R	t-statistics	p-value
Δ^u_{VI}	0.29 ± 0.08	3.433	**0.02**
Δ^u_{VS}	0.26 ± 0.10	2.58	**0.05**
Δ^u_{IS}	0.08 ± 0.08	0.97	0.37

Table 3.1: Relative timing differences Δ in processing information between the three functional units of visual word/object recognition (V), interaction semantics (I) and speech (S), and the T-statistics and p-values of one-sample T-tests used to test the significance of the delays : at the hemodynamic level, using the original BOLD responses (a); and at the neural level, using the reconstructed neural activity functions (b). The delays are measured in terms of the repetition time T_R. The null hypothesis of the one-sample T test is : there is no statistically significant relative latency between regions, e.i. $\Delta = 0$. The alternative hypothesis is: the estimates of relative timing differences is different from zero value, e.i. $\Delta > 0$. The one-sample T-test revealed that, with the 95% confidence, processing information both in the units I and S (marginal case) is significantly delayed relative to V.

In addition to determining the effective connectivity, the same fMRI time series and reconstructed neural functions were used to estimate, through the procedure described in (3.15)-(3.16), the relative timing differences Δ_{VI}, Δ_{VS} and Δ_{IS} in processing information between the visual word/object, interaction-semantics, and speech/phonology systems. We found the following values of the delay, averaged over six observations (subjects): $\Delta^h_{VI} = 0.09 \pm 0.05$, $\Delta^h_{VS} = 0.16 \pm 0.14$ and $\Delta^h_{IS} = 0.03 \pm 0.05$ at the hemodynamic level (see Table 3.1(a)), and $\Delta^u_{VI} = 0.29 \pm 0.08$, $\Delta^u_{VS} = 0.26 \pm 0.10$ and $\Delta^u_{IS} = 0.08 \pm 0.08$ at the neural level (see Table 3.1(b)). After having estimated the mental latency, using each of three pairs of functions, both at the hemodynamic and neural levels, one-sample T-tests were conducted to test the significance of the delay estimates. A one-sample T-test is a statistical method of determining whether the difference between the mean parameters Δ_{VI}, Δ_{VS} and Δ_{IS} and a test value (equal to zero, for the purpose of the present study) is real or merely apparent. Specifically, the null hypothesis to be tested was that there is no significant delay in information flow between the functional units. The alternative hypothesis was that the mean scores of relative timing difference are significantly different from the zero value.

At the hemodynamic level, as supported by the p-values reported in Table 3.1(a), the latency could not be distinguished from zero at the 95% level. In other words, it is impossible to dissect the information flow into its component parts. In contrast to the raw BOLD data, at the neural level, differences in the onset of activation were found to be statistically significant between visual word activation and interaction semantics activation with $p = 0.02$ and a delay of $\Delta_{VI} = (0.29 \pm 0.08) \times T_R =$

950 ms ± 260 ms, as well as between visual word activation and speech production activation (with $p = 0.05$ and a delay of $\Delta_{VS} = (0.26 \pm 0.10) \times T_R = 860\,\text{ms} \pm 330\,\text{ms}$ (see Table 3.1(b)). Latency between semantics and speech could not be discerned. We can put together the results on effective connectivity (Fig. 3.9) with those on the mental chronometry, obtained at the neural level through solving the hemodynamic inverse problem (Table 3.1(b)) and can conclude that, during the object interaction decision task, the sequence of cortical activation leads from the visual word/object system to speech activation via activation in the semantics-associated brain regions, as an intermediary stage, and that the main delay occurs on the pathway leading from the visual word system to interaction semantics. In other words, we found dominance of the mid-dorsal stream in performing a task of reading a word and deciding how to interact with its referent object.

3.5 Discussion

The BOLD response is an indirect indicator of neural activity, and the need for modeling interactions at a neural level motivated the introduction of convolution techniques in fMRI analysis (Friston et al., 1994). In the simplest case, neuronal activity is approximated with the stimulus timing waveform, and can be convolved with an impulse response function parameterized in terms of known basis functions. The validity of the convolution techniques is undermined by the variability of fMRI responses, reported to be different across subjects, from voxel to voxel and from task to task (Handwerker et al., 2004). Another issue is the violation, in many cases, of

the assumption that the underlying neural activity is time-locked to the stimulus. Mathematical models describing the path from neural activity to BOLD is a more advanced way of dealing with the indirect nature of the BOLD mechanism. Furthermore, if the intrinsic biophysical path is specified, extracting the amplitude and timing of underlying neural activity from BOLD responses may potentially provide new information about BOLD dynamics itself.

The purpose of this work was to explore the possibility that solving the hemodynamic inverse problem may allow one to recover some information that was hidden in the translation from the neural level to the relatively sluggish changes in hemodynamics. The proposed techniques were illustrated in a functional model designed to assess the BOLD and neuronal activity involvement in an object interaction decision task. The neuronal activity underlying the BOLD signals were estimated based on the expanded "balloon" model which describes the path from neuronal activity to the BOLD response through the non-linear transitional dynamics of the blood flow-inducing signal, cerebral blood flow, cerebral blood volume and deoxyhemoglobin concentration. The original forward model was transformed into an optimal control problem with the objective of reconstructing the input, the underlying neuronal activity function. Further, effective connectivity between regions associated with visual word/object recognition, interaction semantics, and speech/phonology, and the associated temporal sequencing of information were explored and compared at the hemodynamic and neural levels. Specifically, structural equation modeling was used to estimate causal relationships between the regions, while cross-correlation analysis was applied to determine the relative timing differences between the functional

units involved in performing the task. The causal path pattern at the hemodynamic level was found to be roughly the same as that at the neuronal level. Regarding mental chronometry, using the original fMRI time series failed to reveal significant differences in the relative onset of activation associated with visual word/object, interaction semantics, and speech. Contrary to what was found at the hemodynamic level, at the neuronal level we found a statistically significant delay in activation onset between the visual word system and the semantic system, whereas the delay between the visual word system and speech production system approached significance. Given the strong relationship that exists between nouns (*i.e.*, a word and its referent object) and their associated verbs (*i.e.*, actions associated with the object), it is not surprising that there was only a marginal delay between visual word/object recognition and speech production and there was no significant delay found between interaction semantics and speech/phonology system activation in the present study.

The relative timing differences were found to be measured on the scale of fractions of T_R. In other words, the temporal resolution of the new information obtained at the neural level is higher than the sampling resolution of the original fMRI series. What is the general idea behind the attainment of an increase in temporal resolution? There are a few points worth noting here. First is the inherently approximate nature of the reconstructed neural activity function. The exact solution $u(t)$ is parameterized (see Eq.3.12) with basis functions defined at any time point. Next, from the computational point of view, simulated annealing produces an approximate, although stable, solution from the cluster of admissible solutions associated, in the best scenario, with the global optimum. Finally, the biophysical model itself is a strong

constraint imposed on the definitional domain of the neural input function. It is analogous to the problem to drawing a line through a point, parallel, say, to the ordinate axis (constraint). The matter may be complicated by the intrinsic non-linearity of the biophysical model, but, in fact, there are more relevant issues with the problem of optimal smoothing. There exists a function, called the Hamiltonian, associated with the optimal control problem (Pontryagin, 1962). This function, which describes the state of the system in terms of state variables $\boldsymbol{x}(t) = (x_1(t), x_2(t), x_3(t), x_4(t))^T$ and costate ones $\boldsymbol{p}(t) = (p_1(t), p_2(t), p_3(t), p_4(t))^T$ (Lagrange multipliers), must be minimized over the set of all the possible control functions $u(t)$ (see Vakorin et al. (2006), Chapter 2 of this thesis for details):

$$\mathcal{H}(\boldsymbol{x}(t), u(t), \boldsymbol{p}(t)) = \big(\hat{h}(t) - g(\boldsymbol{x}(t))\big)^2 + \sum_{j=1}^{4} p_j(t) f_j(\boldsymbol{x}(t)) \tag{3.17}$$

The first term of the Hamiltonian, which reflects the discrepancy between the observed and predicted BOLD responses, can be viewed as a minimum square criterion, while the second term puts a constraint on the admissible solution through the transitional dynamics of the physiological variables and their Lagrange multipliers. From this perspective, the hemodynamic inverse problem should be solved through optimal smoothing at the neural level with regularization defined in terms of the biophysical model for transforming the neural activity into BOLD signals.

Another issue is the estimated value *per se* of the activation delay between activation of the visual word/object system and interaction semantics: about 900 ms. From the event-related potential (ERP) literature, a negative-going shift in EEG signal, peaking approximately 400 ms after stimulus onset (N400) is known to be

sensitive to semantic processing (Bentin et al., 1985; Rugg, 1985). Thus, the latency estimated with fMRI data, via solving the inverse problem, is twice as high as that predicted based on EEG data. A factor of 2 is not too bad for a model of the physiology that is as rough as the expanded balloon model. Another reason for the mismatch is the values of the time constants: the flow-inducing signal decay τ_s, the feedback regulatory mechanism τ_f, and the mean transit time of a blood cell in the venous compartment τ_0. These parameters affect the delay of the hemodynamic response, relative to the underlying neural activity, which potentially makes the relative latency parameters more volatile. A better match to known N400 results might be obtained with a better physiological model. For future work, the idea of using tasks that can be decomposed into a sequence of processing stages, together with EEG recordings can guide the investigation of BOLD physiology.

In conclusion, we have demonstrated that solving the hemodynamics inverse problem based on mathematical biophysical models can allow one to extract more information from fMRI signals. In particular, going from the hemodynamic level down to the neural level helped us to explicate more details on mental chronometry. Statistically significant relative timing differences in information processing were found to be smaller than the sampling rate of the raw fMRI data. Mathematical modeling was able to reveal additional neural information from sluggish and indirect fMRI signals, clarifying the neural interpretations of fMRI recordings.

Bibliography

P.A. Bandettini, A. Jesmanowicz, E.C. Wong, and J.S. Hyde. Processing strategies for time-course data sets in functional mri of the human brain. *Magn Reson Med*, 30(2):161–73, 1993.

S Bentin, G. McCarthy, , and C.C. Wood. Event-related potentials associated with semantic priming. *Electroencephalogr Clin Neurophysiol*, 60(4):343–55, 1985.

J. Binder and J. Price. *Handbook of functional neuroimaging of cognition*, chapter Functional neuroimaging of language, pages 187–251. MIT Press, 2001.

R. Borowsky, J. Loehr, C. Kelland Friesen, G. Kraushaar, A. Kingstone, and G.E. Sarty. Modularity and intersection of "what", "where" and "how" processing of visual stimuli: a new method of fMRI localization. *Brain Topogr*, 18(2):67–75, 2005.

R. Borowsky, J. Cummine, W.J. Owen, C.K. Friesen, F. Shih, and G.E. Sarty. fMRI of ventral and dorsal processing streams streams in basic reading processes: Insular sensitivity to phonology. *Brain Topography*, 18(4):233–39, 2006.

J.L. Boxerman, P.A. Bandettini, K.K. Kwong, J.R. Baker, T.L. Davis, B.R. Rosen, and R.M. Weisskoff. The intravascular contribution to fMRI signal change: Monte Carlo modeling and diffusion-weighted studies in vivo. *Magn Reson Med*, 34(1): 4–10, 1995.

C. Buchel and K. Friston. Assessing interactions among neuronal systems using functional neuroimaging. *Neural Netw.*, 13(8-9):871–82, 2000.

C. Buchel and K. Friston. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cereb Cortex*, 7(8):768–78, 1997.

R.B. Buxton, E.C Wong, and L.R. Frank. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn Reson Med*, 39(6): 855–864, 1998.

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

S. Clare, M. Humberstone, J. Hykin, L.D. Blumhardt, R. Bowtell, and P. Morris. Detecting activations in event-related fMRI using analysis of variance. *Magn Reson Med*, 42(6):1117–22, December 1999.

C. de Boor. *A practical guide to Splines*. Springer-Verlag, 1978.

J.B. Demb, R.A. Poldrack, and J.D. Gabrieli. *Functional neuroimaging of word processing in normal and dyslexic readers*, chapter Converging methods for understanding reading and dyslexia, pages 243–304. MIT Press, 1999.

E.A. DeYoe, P. Bandettini, J. Neitz, D. Miller, and P. Winans. Functional magnetic resonance imaging (fMRI) of the human brain. *J Neurosci Methods*, 54(2):171–87, 1994.

E. Formisano and R. Goebel. Tracking cognitive processes with functional MRI mental chronometry. *Curr Opin Neurobiol*, 13(2):174–81, 2003.

K.J. Friston. Bayesian estimation of dynamical systems: an application to fMRI. *Neuroimage*, 16(2):513–30, 2002.

K.J Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Hum Brain Mapp*, 2:56–78, 1994.

K.J. Friston, P. Jezzard, and R. Turner. Analysis of functional MRI time-series. *Hum. Brain Mapp.*, 39:153–71, 1994.

K.J. Friston, A. Mechelli, R. Turner, and C.J. Price. Nonlinear responses in fMRI: the balloon model, volterra kernels, and other hemodynamics. *Neuroimage*, 12(4): 466–477, October 2000.

F. Glover and G.A. Kochenberger. *Handbook of Metaheuristics*. Kluwer Academic Publisher, 2003.

R.L. Grubb, M.E. Phelps, and J.O. Eichling. The effect of vascular changes in PaCO2 on cerebral blood number, blood flow and vascular mean transit time. *Stroke*, 5: 630–9, 1974.

D.A. Handwerker, J.M. Ollinger, and M. D'Esposito. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage*, 21(4):1639–51, 2004.

P.G. Haydon and G. Carmignoto. Astrocyte control of synaptic transmission and neurovascular coupling. *Physiol Rev*, 86(6):1009–31, 2006.

D.J. Heeger and D. Ress. What does fMRI tell us about neuronal activity? *Nat Rev Neurosci*, 3(2):142–51, 2002.

D.J. Heeger, A.C. Huk, W.S. Geisler, and D.J. Albrecht. Spikes versus BOLD: what does the neuroimaging tell us about neural activity. *Nat. Neurosci.*, 3(7):631–3, 2000.

L. Hernandez, D. Badre, D. Noll, and J. Jonides. Temporal sensitivity of event-related fMRI. *Neuroimage*, 17(2):1018–26, 2002.

B. Horwitz. The elusive concept of brain connectivity. *Neuroimage*, 19(2 Pt 1): 466–70, 2003.

B. Horwitz. Relating fmri and pet signals to neural activity by means of large-scale neural models. *Neuroinformatics*, 2(2):251–66, 2004.

B. Horwitz and M.A. Tagamets. Predicting human functional maps with neural net modeling. *Hum Brain Mapp*, 8(2-3):137–42, 1999.

B. Horwitz, M.A. Tagamets, and A.R. McIntosh. Neural modeling, functional brain imaging, and cognition. *Trends Cogn Sci*, 3(3):91–98, 1999.

B. Horwitz, K.J. Friston, and J.G. Taylor. Neural modeling and functional brain imaging: an overview. *Neural netw.*, 13(8-9):829–46, 2000.

F.T. Husain, M.A. Tagamets, S.J. Fromm, A.R. Braun, and B. Horwitz. Relating neuronal dynamics for auditory object processing to neuroimaging activity: a computational modeling and an fMRI study. *Neuroimage*, 21(4):1701–20, 2004.

N.N Kalitkin. *Numerical methods*. Nauka, Moscow, 1978.

N. Kanwisher, P. Downing, R. Epstein, and Z. Kourtzi. *The Handbook on Functional Neuroimaging*, chapter Functional Neuroimaging of Human Visual Recognition, pages 109–152. MIT Press, 2001.

T.Z. Keith. *Multiple regression and beyond*. Pearson Education,Inc., 2006.

R.C. Koehler, D. Gebremedhin, and D.R. Harder. Role of astrocytes in cerebrovascular regulation. *J Appl Physiol*, 100(1):307–17, 2006.

N.K. Logothesis, H. Merkle, M. Augath, T. Trinath, and K. Ugurbil. Neurophisiological investigation of the basis of the fMRI signal. *Nature*, 6843(412):150–157, 2001.

N.K. Logothetis. The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philos Trans R Soc Lond B Biol Sci*, 1424 (357):1003–37, 2002.

J.B. Mandeville, J.J.A. Marota, C. Ayata, G. Zaharchuk, M.A. Moskowitz, B.R. Rosen, and R.M. Weisskoff. Evidence of cerebrovascular postarteriol windkessel with delayed compliance. *J.Cereb.Blood.Flow Metab.*, 19(6):679–89, 1999.

A. Martin. *Handbook of functional neuroimaging of cognition*, chapter Functional neuroimaging of semantic memory, pages 153–87. MIT Press, 2001.

A.R. McIntosh and F. Gonzalez-Lima. Structural equation modeling and its application to network analysis in functional brain imaging. *Hum. Brain Mapp.*, 2(1-2): 2–22, 1994a.

A.R. McIntosh and F. Gonzalez-Lima. Network analysis. *J.Neurosci.*, 14(2):655–66, 1994b.

R.S. Menon and S.G. Kim. Spatial and temporal limits in cognitive neuroimaging with fMRI. *Trends Cogn Sci*, 3(6):207–216, June 1999.

R.S. Menon, D.C. Luknowsky, and J.S. Gati. Mental chronometry using latency-resolved functional MRI. *Proc Natl Acad Sci U S A*, 95(18):10902–7, 1998.

F.M. Miezin, L. Maccotta, J.M. Ollinger, S.E. Petersen, and R.L. Buckner. 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.

A.D. Milner and M.A. Goodale. *The visual brain in action*. Oxford University Press: New York, 1995.

R. Mukamel, H. Gelbard, A. Arieli, U. Hasson, I. Fried, and R. Malach. Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science*, 309(5736):951–4, 2005.

D.G. Nair. About being BOLD. *Brain Res Brain Res Rev*, 50(2):229–43, 2005.

S Ogawa, T.M. Lee, A.R. Kay, and D.W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S A*, 87(24): 9868–72, 1990.

S. Ogawa, R.S. Menon, D.W. Tank, S.G. Kim, H. Merkle, J.M. Ellerman, and K. Ugurbil. Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging: a comparison of signal characteristics with a biophysical model. *Biophys.J.*, 3(64):803–812, 1993.

E.D. Palmer, T.T. Brown, S.E. Petersen, and B.L. Schlaggar. Investigation of the functional neuroanatomy of single word reading and its development. *Scientific studies of language*, 8:203–23, 2004.

L.S. Pontryagin. *Mathematical Theory of Optimal Processes*. John Wiley & Sons, 1962.

M.I. Posner. *Chronometric explorations of mind*. Hillsdale, 1978.

G. Rees, K. Friston, and C. Koch. A direct quantative relationship between the functional properties of human and macaque V5. *Nat. Neurosci.*, 3(7):719–23, 2000.

J.J. Riera, J. Watanabe, I. Kazuki, M. Naoki, E. Aubert, Ozaki.T., and R. Kawashima. A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signals. *Neuroimage*, 21(2):547–67, 2004.

M.D. Rugg. The effects of semantic priming and work repetition on event-related potentials. *Psychophysiology*, 22(6):642–7, 1985.

Z.S. Saad, E.A. DeYoe, and K.M. Ropella. Estimation of fmri response delays. *Neuroimage*, 18(2):494–504, 2003.

G.E Sarty. *Computing brain activity maps from fMRI time-series images*. Cambridge University Press, 2007.

G.E. Sarty and R. Borowsky. Functional MRI activation maps from empirically defined curve fitting. *Concepts in Magnetic Resonance Part B*, 24B(1):46–55, 2005.

T.J. Sejnowski, C. Koch, and P.S. Churchland. Computational neuroscience. *Science*, 241(4871):1299–306, 1988.

A. Shmuel, M. Augath, A. Oeltermann, and N.K. Logothetis. Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. *Nat Neurosci*, 9(4):569–77, 2006.

M.A. Tagamets and B. Horwitz. Integrating electrophysiological and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain imaging study. *Cereb Cortex*, 8(4):310–20, 1998.

T. Takano, G.F. Tian, W. Peng, N. Lou, W. Libionka, X. Han, and M. Nedergaard. Astrocyte-mediated control of cerebral blood flow. *Nat Neurosci*, 9(2):260–7, 2006.

V.A. Vakorin, O.O. Krakovska, R. Borowsky, and G.E. Sarty. Inferring neural activity from BOLD signals through nonlinear optimization. *Neuroimage, under review*, 2006.

X. Wan, J. Riera, K. Iwata, M. Takahashi, T. Wakabayashi, and R. Kawashima. The neural basis of the hemodynamic response nonlinearity in human primary visual cortex: Implications for neurovascular coupling mechanism. *Neuroimage*, 32(2): 616–25, 2006.

Y. Zheng, J. Martindale, D. Johnston, M. Jones, J. Berwick, and Mayhew J. A model of the hemodynamic response and oxygen delivery to brain. *Neuroimage*, 16(3):617–37, July 2002.

M. Zonta, M.C. Angulo, S. Gobbo, B. Rosengarten, K.A. Hossmann, T. Pozzan, and G. Carmignoto. Neuron-to-astrocyte signaling is central to the dynamic control of brain microcirculation. *Nat Neurosci*, 6(1):43–50, 2003.

Chapter 4

Characterizing the functional MRI response using Tikhonov regularization

4.1 Preliminaries

This chapter is based on the manuscript titled "Characterizing the fMRI response using Tikhonov regularization" (2006) by Vakorin,V.A., Borowsky,R. and Sarty,G.E., submitted to *Statistics in Medicine*. As it was shown in Chapter 3, solving the hemodynamic inverse problem, *i.e.* the problem of extracting neural dynamics, can reveal information that was not available at the level of fMRI BOLD data. Specifically, at the neural level, we were able to determine statistically significant latencies in activation between functional units in a model of an object interaction decision task. Mathematically, solving the hemodynamic inverse problem in Chapter 3 was equivalent to smoothing at the neural level, accompanied with the idea of signal averaging. In this chapter, a similar problem is considered at the level of BOLD fMRI time series *per se*. This work is an attempt to probe the issue of what amount of smoothing should be considered optimal for fMRI signals when assessing activation. This is motivated by reasoning that strong smoothing can eliminate relevant temporal characteristics while insufficient smoothing may leave much noise in hemodynamic

responses.

Specifically, the problem of evaluating an averaged functional magnetic resonance (fMRI) response for repeated block design experiments was considered within a semiparametric regression model with autocorrelated residuals. We applied functional data analysis (FDA) techniques that use a least squares fitting of B-spline expansions with Tikhonov regularization. To deal with the noise autocorrelation, we proposed a regularization parameter selection method based on the idea of combining temporal smoothing with residual whitening. A criterion based on a generalized χ^2 test of the residuals for white noise was compared to a generalized cross validation scheme. We evaluated and compared the performance of the two criteria, based on their effect on the quality of the fMRI response. We found that the regularization parameter can be tuned to improve the noise autocorrelation structure but the whitening criterion provides too much smoothing when compared to the cross-validation criterion.

The ultimate goal of the proposed smoothing techniques is to facilitate the extraction of temporal features in the hemodynamic response for further analysis. In particular, these FDA methods allow us to compute derivatives and integrals of the fMRI signal so that fMRI data may be correlated with behavioral and physiological models. For example, positive and negative hemodynamic responses may be easily and robustly identified on the basis of the first derivative at an early time point in the response. Ultimately these methods allowed us to verify previously reported correlations between the hemodynamic response and the behavioral measures of accuracy and reaction time, showing the potential to recover new information from

fMRI data (see Chapter 5).

4.2 Introduction

Functional magnetic resonance imaging (fMRI) is a non-invasive technique used along with specific cognitive tasks to measure the dynamics of the hemodynamic response of the brain in reaction to that task. The two primary objectives for fMRI analysis are the segmentation of activated brain regions and the characterization of the temporal response. A variety of techniques have been proposed to detect activation patterns from fMRI time series. Traditionally, these techniques are based on calculating correlations of intensity time series with an *a priori* given reference function representing a convolution of the stimulus with an assumed hemodynamic impulse response function. Here we concentrate on the problem of characterizing the temporal response.

The simplest fMRI experimental design is the block design consisting of periodically repeated blocks which in turn, are composed of image volumes scanned during a rest condition, followed by volumes scanned during a stimulus condition. More advanced experimental designs involve the event-related design (Friston et al., 1998). Event-related data can be analyzed by calculating two-condition t-tests that compare a rest baseline condition with an experimental one. This t-test is equivalent to a correlation between the observed time-series and a stepwise function that crudely approximates the expected hemodynamic response. More refined model-oriented methods calculate correlations with more realistic hemodynamic response functions,

such as gamma functions (Frackowiak et al., 1997), but the reference function remains the same for every voxel.

A correlation-based method, termed the BOLDfold (Sarty and Borowsky, 2005), was proposed for repeated block design experiments in which the comparison function is assumed to be different for every voxel and can be defined empirically through periodic data folding. Mathematically, the BOLDfold approach can be shown to be equivalent to an analysis of variance approach (Clare et al., 1999). In BOLDfold, the activation maps are calculated voxelwise based on the correlations between the intensity time course and its repeated mean time course. Furthermore, the time series themselves for the voxels found to be activated are of special interest since the temporal characteristics of the fMRI response are thought to be informative with regard to mental activity (Menon et al., 1998). Specifically, the fMRI onset latency, defined as the time at which the hemodynamic response intensity reaches its maximum value, has been shown to correlate with reaction time and stimulus presentation time (Menon et al., 1998). Other studies suggest that the full-width-at-half-maximum (FWHM) parameter is better correlated to the reaction time (Liu et al., 2004).

Because of the discrete and noisy nature of the fMRI response, evaluating temporal parameters of the hemodynamic response inexorably requires some kind of smoothing. In this paper we focus on functional data analysis (FDA) methods (Ramsay and Silverman, 2002) for an optimal smoothing of the averaged fMRI response. Mathematically the selection criterion was chosen to minimize the energy of approximation $\mathcal{F}$ by the parameters $\boldsymbol{\beta}$ that characterize the fMRI response according to

the scheme

$$\mathcal{F}(\boldsymbol{\beta}) = \mathcal{F}_1(\boldsymbol{\beta}) + \lambda\, \mathcal{F}_2(\boldsymbol{\beta})$$

where the term $\mathcal{F}_1(\boldsymbol{\beta})$ is the quadratic error, and the regularization term $\mathcal{F}_2(\boldsymbol{\beta})$ represents a roughness of the approximation, a property quantified by the integrated sum of the local variations of the derived fMRI response. The regularization parameter λ controls the amount of applied smoothing, and consequently the temporal properties of the fMRI response. Furthermore, the autocorrelation structure of residuals can be improved through tuning the parameter λ.

Violation of the assumption of independent residuals in fMRI analysis has been a problem for obtaining correct estimations of and inferences about the evoked hemodynamic response. Functional MRI noise has many components, both physiological and instrumental, which is exhibited in autocorrelated residuals in the general linear models (GLM) used to analyze the response (Friston et al., 1995a). To deal with the noise autocorrelation, temporal filtering (Friston et al., 1995b; Worsley and Friston, 1995) and residual whitening (Bullmore et al., 1996a) have been proposed in the literature. Whitening is the most efficient approach to parameter estimation provided that there is no discrepancy between the assumed and actual autocorrelations. Since in practice it is quite challenging to obtain an accurate estimation of the true autocorrelations, whitening makes the analysis sensitive to bias, which affects statistical inferences. It has been argued in Friston et al. (2000), therefore, that it is preferable to smooth in order to minimize the bias rather than whiten the residuals.

Spline smoothing which is asymptotically equivalent to a kernel smoothing in

the kernel density estimation setting (Eubank, 1988) provides a powerful tool for obtaining insights into the data. However, the effective use of these techniques involves selecting the smoothing parameter. In attempting to find an optimal amount of smoothing as governed by the regularization parameter λ, Clare et al. (1999) applied a generalized cross-validation (GCV) scheme in a spline smoothing of the fMRI time series. GCV can be viewed as a criterion to avoid the overfitting of data in the bias/variance dilemma representing a trade-off between a gain in bias reduction and an increase in variance. Many examples of applying GCV for automatic smoothing can be found in a number of monographs (Ramsay and Silverman, 2002; Eubank, 1988; Wahba, 1990). In addition, the issue of using GCV as a possible candidate for the best criterion for automatic regularization has been recently addressed in a number of medical applications: ultrasonic pulse-echo imaging (Lavarello et al., 2006), electrical impedance tomography (Graham and Adler, 2006), perfusion MRI (Sourbron et al., 2004) and positron emission tomography (Lu et al., 1998).

As an alternative to the GCV scheme for functional MRI data analysis, we proposed and examined a criterion which combines a smoothing scheme with the idea of minimizing noise autocorrelation. Specifically, a generalized χ^2 test of residuals for white noise was employed to find the optimal value of the regularization parameter λ. The cross-validation and whitening criteria were compared with each other based on the amount of smoothing they provide for the time course of activated voxels.

In summary, we present here a smoothing scheme for evaluating the mean fMRI response using the FDA techniques. We aimed to (i) propose and justify techniques that facilitate calculations of temporal characteristics of the fMRI response and (ii)

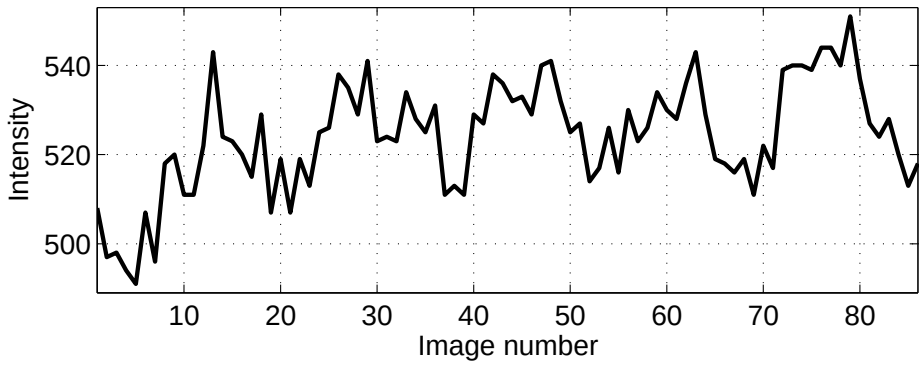
refine criteria for selecting the optimal degree of smoothing. With the smoothed hemodynamic response, maps of certain temporal features such as the FWHM and the onset time, can be easily constructed and used for further analysis of mental operations in perceptual and cognitive tasks.

4.3 Methods

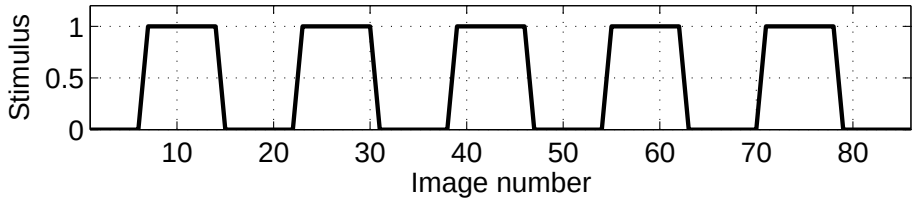
4.3.1 Smoothing averaged responses

Signal averaging is widely known as a technique that can be used to increase signal-to-noise ratio of fMRI signals (DeYoe et al., 1994). It is employed in the BOLDfold approach to produce a reference function individual for every voxel. Mathematically, the fMRI data from repeated block experiments are folded over and averaged to yield the hemodynamic response function. Specifically, let j index the P repeated imaging blocks (periods), and let i index the T time point samples $\{t_i\}_{i=1}^{T}$ for each block. An example fMRI time series obtained in an experiment with 5 blocks is provided in Fig. 4.1(a), while Fig. 4.1(b) illustrates the corresponding paradigm function, showing the periods of stimulus and rest. Also, let $\mathbf{y} = \{y_s\}_{s=1}^{N}$ denote the observed time series where $N = P\,T$, and let $\mathbf{h} = \{h_i\}_{i=1}^{T}$ denote the averaged, folded data. For every t_i we can find h_i by picking up the values of the fMRI response $\mathbf{y}$ corresponding to the i^{th} element in each block, namely $\{y_{(i+j\,T)}\}_{j=1}^{P}$, and computing the mean vector $\boldsymbol{h}$ with components

$$h_i = \frac{1}{P}\sum_{j=1}^{P} y_{(i+j\,T)} \tag{4.1}$$



(a) fMRI time series.



(b) Paradigm function.

Figure 4.1: An example of fMRI time series obtained in an experiment with 5 blocks and the corresponding paradigm function.

In the BOLDfold method, the time series for every voxel is correlated with its own averaged repeated time series. Specifically, the averaged, folded signal $\boldsymbol{h}$ can be repeated P times producing the time series defined by the vector

$$\tilde{\mathbf{h}}^T = \mathbf{1}^T \bigotimes \mathbf{h}^T = \underbrace{\left[\begin{array}{ccc} \mathbf{h}^T & \cdots & \mathbf{h}^T \end{array} \right]}_{P \text{ times}}$$

where $\mathbf{1}$ is the vector of the length P, and the symbol $\bigotimes$ denotes the Kronecker product. Then, this newly built, unfolded signal $\tilde{\mathbf{h}}$ spanning the whole experiment can be correlated with the original time series $\boldsymbol{y}$. The correlation coefficients r contains only empirically obtained data terms. High correlations between $\mathbf{y}$ and $\tilde{\mathbf{h}}$

reflect the activation of a specific brain area in reaction to the given stimulus. As mentioned in the introduction, it has been shown that the BOLDfold method for computing activation maps is mathematically equivalent to an analysis of variance (ANOVA) method that produces F statistic maps (Clare et al., 1999).

A more general way to compute the averaged fMRI signal is to apply the general linear model (GLM) for the time series data as follows. Let

$$\mathbf{y} = \mathbf{B}\mathbf{h} + \mathbf{L}\boldsymbol{\zeta} \tag{4.2}$$

where $\mathbf{B}$ is the design matrix associated with modeling the averaged response $\boldsymbol{h}$, $\mathbf{L}$ is an unknown convolution matrix which describes noise autocorrelation, and $\boldsymbol{\zeta} = \{\zeta_i\}_{i=1}^{N}$ is a vector of zero mean, uncorrelated random errors with common variance σ_ζ^2, $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \sigma_\zeta^2 \mathbf{I}_N)$. Let the $(N \times T)$ design matrix $\mathbf{B}$ be defined as the Kronecker product of the P-dimensional vector $\mathbf{1}$ of ones and the $(T \times T)$ identity matrix $\mathbf{I}_T$, namely

$$\mathbf{B}^T = \mathbf{1}^T \bigotimes \mathbf{I}_T = \underbrace{\left[\, \mathbf{I}_T \quad \cdots \quad \mathbf{I}_T \,\right]}_{P \text{ times}}$$

This block-based form of the design matrix $\boldsymbol{B}$ is designed to parameterize the shape of the average hemodynamic response over blocks in a blocked design. Other variations may be possible for event-related experimental designs, for example, those aimed at characterizing the impulse response function. However, we will not pursue those variations here.

If we assume the simplification $\mathbf{L} = \mathbf{I}_N$, then the ordinary least squared (OLS) estimate $\hat{\mathbf{h}} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{y}$ gives exactly the averaged repeated hemodynamic re-

sponse calculated in (4.1). The model (4.2) can be expanded by implementing a $(N \times M)$ nuisance matrix $\mathbf{D}$ that accounts for effects of no interest, such as non-specific signal drift, which are modeled in general by a vector $\boldsymbol{\mu}$ of length M; so the expanded model is

$$\mathbf{y} = \mathbf{B}\mathbf{h} + \mathbf{D}\boldsymbol{\mu} + \mathbf{L}\boldsymbol{\zeta} = [\,\mathbf{B}\;\mathbf{D}\,] \begin{bmatrix} \mathbf{h} \\ \boldsymbol{\mu} \end{bmatrix} + \mathbf{L}\boldsymbol{\zeta}. \tag{4.3}$$

Our objective is to fit the discrete averaged observations $h_i, i = 1, \ldots, T$ to a function $h(t)$ with the smoothing model $h_i = h(t_i) + \eta_i$, where the elements η_i are uncorrelated random errors $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0},\, \sigma_\eta^2\, \mathbf{I}_N)$. We will use a basis function expansion for $h(t)$ of the form

$$h(t) = \sum_{j=1}^{K} c_j\, \phi_j\,(t) \tag{4.4}$$

where $\{\phi_j\}_{j=1}^K$ is a set of known basis functions. To test the proposed techniques we will use B-splines (de Boor, 1978) for spline smoothing. B-splines were chosen because of their stable numerical properties, their continuous derivatives at the joining points and their local support which prevents an observation from affecting the entire shape of the hemodynamic response. Now let the K-dimensional vector $\boldsymbol{c}$ contain the coefficients $\{c_j\}_{j=1}^K$, while the $(T \times K)$ matrix $\boldsymbol{\Phi} = \{\phi_{j\,i}\}$ contains the values $\phi_j(t_i)$. Then

$$\mathbf{h} = \boldsymbol{\Phi}\mathbf{c} + \boldsymbol{\eta} \tag{4.5}$$

and the model (4.3) becomes

$$\mathbf{y} = [\,\mathbf{B}\quad\mathbf{D}\,] \begin{bmatrix} \boldsymbol{\Phi}\boldsymbol{c} + \boldsymbol{\eta} \\ \boldsymbol{\mu} \end{bmatrix} + \mathbf{L}\boldsymbol{\zeta} = [\,\mathbf{B}\boldsymbol{\Phi}\quad\mathbf{D}\,] \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{\mu} \end{bmatrix} + \boldsymbol{B}\boldsymbol{\eta} + \mathbf{L}\boldsymbol{\zeta} \tag{4.6}$$

or

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{K}\boldsymbol{\varepsilon} \tag{4.7}$$

where $\mathbf{B\Phi}$ and $\mathbf{D}$ are partitions of the design matrix $\mathbf{X} = [\mathbf{B\Phi} \;\; \mathbf{D}]$, and $\boldsymbol{\beta} = [\boldsymbol{c}^T \; \boldsymbol{\mu}^T]^T$ is a column vector of parameters for the effects modeled by each column of the design matrix. The matrix $\boldsymbol{K}$ represents the effective convolution matrix which describes the new autocorrelation, and $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_N)$, such that $\boldsymbol{K}\boldsymbol{\varepsilon} = \boldsymbol{B}\boldsymbol{\eta} + \mathbf{L}\boldsymbol{\zeta}$.

Let $\mathbf{S}$ be the matrix of a temporal filtering transformation applied to the model (4.7) to give

$$\boldsymbol{S}\boldsymbol{y} = \boldsymbol{S}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{S}\boldsymbol{K}\boldsymbol{\varepsilon} \tag{4.8}$$

In order to whiten the data, we need to know the intrinsic autocorrelation structure $\boldsymbol{K}$. If $\boldsymbol{K}$ is known, choosing $\boldsymbol{S} = \boldsymbol{K}^{-1}$ provides minimum variance, unbiased OLS estimates $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{S}\boldsymbol{y}$. Practically, $\boldsymbol{K}$ is never known, which means that the true intrinsic correlations $\mathbf{K}$ need to be estimated. In this case $\boldsymbol{S} \neq \boldsymbol{K}^{-1}$, and the assumed autocorrelation $\boldsymbol{V}_a = \mathbf{K}_a^T\mathbf{K}_a$ differs from the true autocorrelation $\boldsymbol{V} = \mathbf{K}^T\mathbf{K}$. That leads to a bias between the true contrast variance $\mathrm{Var}(\mathbf{a}^T\hat{\boldsymbol{\beta}})$ and the expectation of its estimator $E\big[\mathrm{Var}(\mathbf{a}^T\hat{\boldsymbol{\beta}})\big]$ (here $\mathbf{a}$ is a vector of contrast weights). Friston et al. (2000) found that a modified $1/f$ model (Zarahn et al., 1997) and an autoregressive AR(1) model (Bullmore et al., 1996b) fail to capture long-range correlations, leading to unacceptably reduced efficiency and inflated contrast variance. An alternative solution involves conditioning the correlations by smoothing. In that case, even if $\boldsymbol{V}_a$ is not known, smoothing provides an acceptable level of bias (Worsley and Friston, 1995). This justifies the implementation of the basic function expansion

of $h(t)$ in (5.2), which serves to temporally filter the noisy $\boldsymbol{h}$.

Now we will show how including the smoothing matrix $\mathbf{S}$ corresponds to an objective functional $\mathcal{F}$ based on the least squares criterion including a regularization term. In the context of this paper, Tikhonov regularization can be used to impose constraints on derivatives of the smoothed averaged fMRI response $h(t)$ via the functional

$$\mathcal{F} = \| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \|^2 + \lambda \int_{t_1}^{t_T} |\boldsymbol{\mathcal{L}}\, h(t)|^2 dt \tag{4.9}$$

where $\boldsymbol{\mathcal{L}}$ is a linear differential operator and λ is the regularization parameter. An estimator of β can be obtained from

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \mathcal{F}. \tag{4.10}$$

The parameter $\lambda \geq 0$ can be used to control the trade-off between smoothness and closeness of fit. Strong smoothing but minimal bending is obtained when $\lambda \to \infty$. Bending becomes more significant with $\lambda \to 0$. To choose λ is a fundamental problem, known as the bias/variance dilemma (see, for example, Chanmond (2003)). In the context of Bayesian framework modeling and the autocorrelation structure of the residuals, λ can be interpreted in terms of a noise level (Goutte et al., 2000).

With (5.2), the second term of the cost functional (4.10) becomes

$$\begin{aligned}
\lambda \int_{t_1}^{t_T} |\boldsymbol{\mathcal{L}}\, h(t)|^2 dt &= \lambda \int_{t_1}^{t_T} |\boldsymbol{\mathcal{L}} \sum_{j=1}^{K} c_j\, \phi_j\,(t)|^2 dt \\
&= \lambda \int_{t_1}^{t_T} \left(\sum_{i,j=1}^{K} c_i\, \boldsymbol{\mathcal{L}}\, \phi_i\,(t)\ \boldsymbol{\mathcal{L}}\, \phi_j\,(t)\ c_j \right) dt \\
&= \lambda \sum_{i,j=1}^{K} c_i \left(\int_{t_1}^{t_T} \boldsymbol{\mathcal{L}}\, \phi_i\,(t)\ \boldsymbol{\mathcal{L}}\, \phi_j\,(t)\, dt \right) c_j \\
&= \lambda\, c^{\top} \mathbf{R}\, c \qquad (4.11)
\end{aligned}$$

where $\boldsymbol{R} = \{r_{ij}\}_{i,j=1}^{K}$ is the $(K \times K)$ matrix such that $r_{ij} = \int_{t_1}^{t_T} \boldsymbol{\mathcal{L}}\, \phi_i\,(t)\ \boldsymbol{\mathcal{L}}\, \phi_j\,(t)\, dt$. Thus the functional (4.9) may be expressed in matrix terms as

$$\mathcal{F} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda\, c^T \mathbf{R}\, c. \qquad (4.12)$$

Let $\boldsymbol{J}$ be defined as the $(K \times (K + M))$ matrix $\mathbf{J} = [\, \mathbf{I}_K\ \mathbf{O}\,]$ where $\mathbf{O}$ is the $(K \times M)$ matrix with all the elements equal to zero. It is easy to show that $\boldsymbol{c}^T\, \mathbf{R}\, \boldsymbol{c} = \boldsymbol{\beta}^T \mathbf{J}^T \mathbf{R}\, \mathbf{J} \boldsymbol{\beta}$.

Defining the $(K + M) \times (K + M)$ matrix $\boldsymbol{\Sigma}$ as $\boldsymbol{\Sigma} = \mathbf{J}^T\, \mathbf{R}\, \mathbf{J}$, the regularization term in (4.12) becomes $\lambda\, \boldsymbol{\beta}^T \boldsymbol{\Sigma}\, \boldsymbol{\beta}$. It can be shown that the solution to the problem (4.10) is

$$\hat{\boldsymbol{\beta}}_\lambda = \left(\mathbf{X}^{\top} \mathbf{X} + \lambda\, \boldsymbol{\Sigma} \right)^{-1} \mathbf{X}^T \boldsymbol{y} \qquad (4.13)$$

where we use the subindex λ to underline the dependence of the estimated parameters $\hat{\boldsymbol{\beta}} = \left[\, \hat{\boldsymbol{c}}^T\, \hat{\boldsymbol{\mu}}^T \,\right]^T$ on the regularization parameter λ. We note that for $\lambda \neq 0$ the matrix $(\boldsymbol{X}^T \boldsymbol{X} + \lambda \boldsymbol{\Sigma})$ is invertible by construction. In the case of $\lambda = 0$, multicollinearity is present only when the number of basis functions used to approximate $\boldsymbol{h}$ exceeds the number of points wherein $\boldsymbol{h}$ is considered, *i.e.* if $K > T$. Then, with

$\hat{\boldsymbol{c}}$ evaluated through (4.13), (5.2) provides a smoothed representation of the averaged fMRI response $h(t)$. Furthermore, the smoothing matrix $\boldsymbol{S}$ used in (4.8) for the Tikhonov regularization is $\boldsymbol{S}_\lambda = \boldsymbol{X}\left(\mathbf{X}^T\mathbf{X} + \lambda\,\boldsymbol{\Sigma}\right)^{-1}\mathbf{X}^T$. The matrix $\boldsymbol{S}_\lambda$ determines df_{fit}, the degrees of freedom of the fit corresponding to the spline smoothing, namely $df_{\text{fit}}(\lambda) = \text{tr}\boldsymbol{S}_\lambda$ (Ramsay and Silverman, 2002) with an interpretation as the equivalent number of parameters. We report monotonically decreasing relationships between $df_{\text{fit}}(\lambda)$ and λ. The residuals can be calculated as

$$\hat{\boldsymbol{\varepsilon}}_\lambda = \boldsymbol{y} - \hat{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{S}_\lambda)\boldsymbol{y}. \tag{4.14}$$

4.3.2 Criteria for choosing the regularization parameter

The variational problem (4.10) has so far been considered without mention of how to choose the parameter λ automatically. A variety of criteria for choosing λ have been proposed in the literature (Eubank, 1988; Wahba, 1990). A classical criterion is "cross-validation" which relies on splitting the data into estimation and validation parts. The simplified leave-one-out case is based on the idea of deleting an observation and approximating the data with the rest of the observations. Repeating the procedure of deleting and fitting for all observations makes it possible to measure the quality of approximation through the cumulative quadratic error. The quadratic error criterion can be refined (Craven and Wahba, 1979) to produce the generalized cross validation (GCV) measure as

$$GCV(\lambda) = \frac{1}{N}\frac{\hat{\boldsymbol{\varepsilon}}_\lambda^T\hat{\boldsymbol{\varepsilon}}_\lambda}{\left(1 - \frac{1}{N}\text{tr}\boldsymbol{S}_\lambda\right)^2}. \tag{4.15}$$

Since we are interested in optimizing the correlation structure of the residuals, we will also consider a criterion for choosing λ through optimizing the value of a white noise test for the residuals. The generalized χ^2-test is based on the idea that a stationary process is white noise if and only if its autocorrelation function $\rho(k)_\lambda$ is zero for lag one and above. This leads to the test statistic $N\sum_{k=1}^{m}\hat{\rho}_\lambda(k)^2$ where $\hat{\rho}_\lambda$ is an estimated autocorrelation function for the process defined by the time series $\hat{\boldsymbol{\varepsilon}}_\lambda$ in (4.14). However, possible high deviations of this statistic from the corresponding χ^2 distribution (Davies et al., 1977) together with suspiciously low values reported in Ljung and Box (1978) for this statistic led to a corrected test. A better approximation was shown to be achieved (Ljung and Box, 1978) by using

$$T(\lambda) = N(N+2)\sum_{k=1}^{m}\frac{\hat{\rho}_\lambda(k)^2}{N-k} \tag{4.16}$$

for a given number of parameters m.

So, in order to choose an optimal λ we have to solve optimization (minimization) problems with the objective functions defined using either (4.15) or (4.16). The optimal value of the parameter λ determines the optimal amount of smoothing provided by the smoothing matrix $\mathbf{S}$ in (4.8). At the same time, the parameter λ conditions the estimates for $\boldsymbol{\beta}$, and consequently, the estimates for the parameters $\boldsymbol{c}$ of the smoothed representation (5.2) of $\mathbf{h}$.

4.4 Application

4.4.1 fMRI experiment

A functional imaging experiment involving a finger-touching motor task was performed with a 1.5 T Siemens Symphony scanner. The experiment was similar to one used in a study of the role of the left hemisphere in motor control of touch (Borowsky et al., 2002). A gradient echo T_2^* single-shot EPI sequence ($T_E = 55$ms and $T_R = 1600$ms) was used with fat saturation pulses. The 12 slices acquired had a square field of view of 250 mm and were 8-mm thick with an interslice gap of 2 mm. The original 64×64 data matrices were Fourier reconstructed to 128×128 images. The stimulus was presented for 8 volumes followed by 8 volumes of rest (see Fig. 4.1(b)), leading to a block consisting of 16 volumes ($T = 16$). The block pattern was repeated 5 times ($P = 5$). The first 6 image volumes, collected before the beginning of the first block to allow the spins to reach a steady state, were discarded prior to analysis. So the complete dataset consisted of 86 image volumes of which 80 were used in the analysis.

4.4.2 Analysis

The differential operator in (4.9) was defined the second derivative $\mathcal{L} = \frac{\mathrm{d}^2}{\mathrm{d}^2 t}$. A result found in de Boor (1978) states that the curve $h(t)$ that minimizes (4.9) should be a cubic spline with knots at the data points t_i. We used an order four B-spline system $\{\phi_j\}_{j=1}^{K}$ (de Boor, 1978) for the basis function expansion (5.2) of the mean response

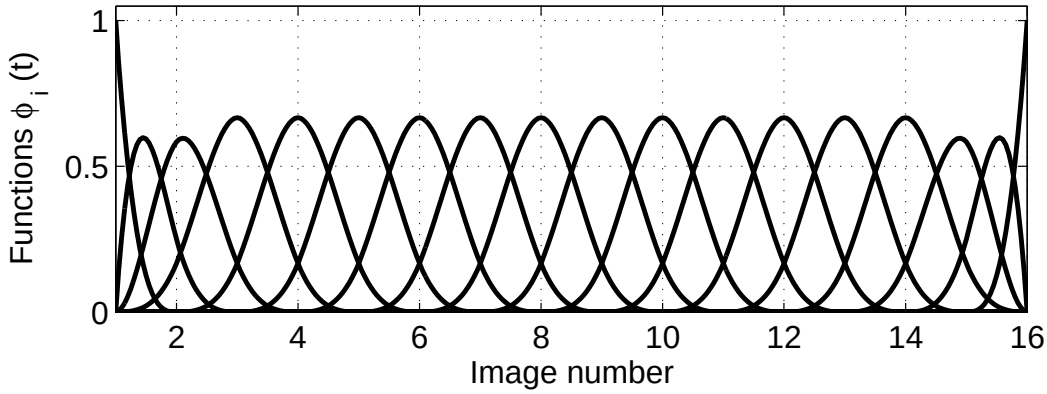


Figure 4.2: A set of 18 B-spline functions defined on the interval $1 \leq t \leq 16$. A B-spline is a spline function which has minimal support with respect to a given degree, smoothness, and domain partition. Each basis function $\phi_j(t)$ illustrated is a piecewise polynomial of the third degree. Note that $\sum_{j=1}^{18} \phi_j(t) = 1$ for any $t \in [1, 16]$.

$h(t)$. Each basis function $\phi_j(t)$ is a spline function defined by an order and a knot sequence. Taking into account the relation between the order of the spline ($R = 4$), number of knots ($T = 16$) and the number of basis functions, using cubic splines implies that we have $(T + 2)$ basis functions. Since we fit T points with a linear combination of $(T + 2)$ basis functions, regularization is necessary, which excludes the case $\lambda = 0$. Fig. 4.2 illustrates the basis consisting of 18 B-spline functions of order 4 (cubic B-splines) defined on the interval $1 \leq t \leq 16$. The three basis functions on both sides of the interval are different. As we move from the center to the left (or to the right in reverse order), the basis gradually changes in smoothness, from being twice-differentiable to being discontinuous. As commented in Ramsay and Silverman (2002), losing differentiability at the boundaries makes some sense because no information is available beyond the interval. However, taking into account the normalized nature of the B-splines basis in the form of $\sum_{j=1}^{K} \phi_j(t) = 1$ for any $t \in [1, T]$ including the edge regions and shape preserving property of B-splines

(Carnicer and Pena, 1994), the boundary effect appears to be not significant.

We note that dimension of the parameter λ depends on the choice of the differential operator $\mathcal{L}$. In the case of $\mathcal{L} = \frac{\mathrm{d}^2}{\mathrm{d}^2 t}$, the dimension of λ becomes $[s^{-2}]$, or $[T_R^{-2}]$ putting it in terms of an image sequence when the frame acquisition rate is defined by T_R. This, in turn, means that the dimension of $1/\sqrt{\lambda}$ is essentially the bandwidth in kernel estimating schemes (Eubank, 1988) and is measured in the units of T_R.

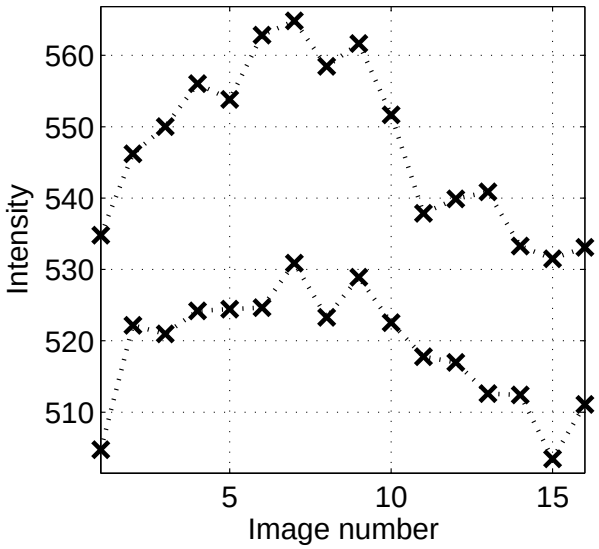
To compare the two criteria for λ represented by the objective functions $GCV(\lambda)$ and $T(\lambda)$, we considered a time series of voxels as activated if ordinary BOLDfold analysis showed it as being active. This was determined as follows. First, for every voxel in the dataset with a minimum observed intensity above a background cut-off of 200 grey-scale units, the mean hemodynamic responses $\mathbf{h}$ were calculated through the simple averaging procedure (4.1). Then, the response $\mathbf{h}$ was repeated 5 times to produce the repeated BOLDfold function $\tilde{\mathbf{h}}$ spanning the entire period of the experiment. The activation maps based on the correlation between the function $\tilde{\mathbf{h}}$ and the original time course $\mathbf{y}$ were computed. The correlation range $r \geq 0.7$ was used to define a set $\mathcal{A}$ containing all the activated voxels. It should be noted that the set $\mathcal{A}$ contained both a subset $\mathcal{A}^+ \subset \mathcal{A}$ of voxels showing a positive, conventional increase in fMRI signal in response to a stimulus and a subset $\mathcal{A}^- \subset \mathcal{A}$ of voxels with a negative response (Harel et al., 2002; Shmuel et al., 2006). Examples of the mean fMRI signal from the sets $\mathcal{A}^+$ and $\mathcal{A}^-$ are shown in Fig. 4.3(a) and 4.3(b) respectively. $\mathcal{A}^+$ and $\mathcal{A}^-$ were separated based on the first derivative maps of the smoothed hemodynamic response $h(t)$ evaluated at the point t_1. To compute the maps of $\frac{\mathrm{d}h(t)}{\mathrm{d}t}|_{t=t_1}$, $h(t)$ was smoothed with the B-splines for $\lambda = 1$. Only those voxels

demonstrating conventional positive response were considered for further analysis. The conventional increase in the intensity of the hemodynamic signal evoked by increased neuronal activity is thought to be due to uncoupling between local cerebral blood flow and the cerebral rate of oxygen consumption (Buxton et al., 1998). We focused on interpreting the positive BOLD responses because there is some controversy in the literature over the nature of the $\mathcal{A}^-$ type of responses (Shmuel et al., 2006).

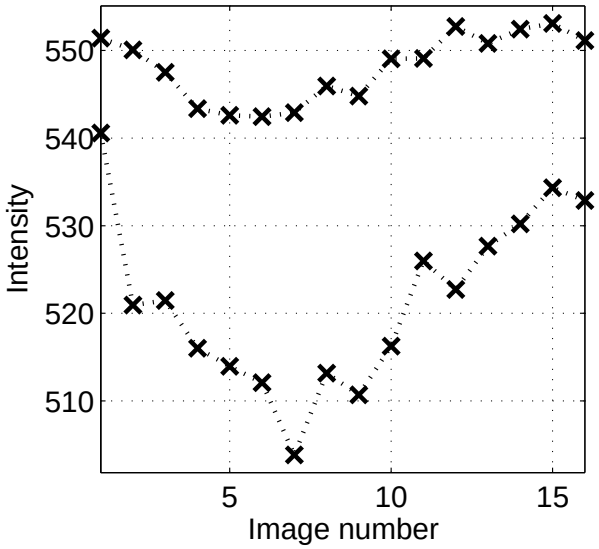
For every voxel from the set $\mathcal{A}^+$, the optimal amount of smoothing was evaluated using the criteria (4.15) and (4.16). Specifically, the functions $GCV(\lambda)$ and $T(\lambda)$ were minimized over λ on the interval $1 \leq \lambda \leq 10^4$. Fig. 4.4 gives an example of the functions $GCV(\lambda)$ and $T(\lambda)$ for a voxel found activated in the ordinary BOLDfold analysis. As can be seen from Fig. 4.4(b), the smoothing parameter λ indeed affects the noise autocorrelation structure in the GLM model.

The distribution of the optimal λ_{GCV} for the voxels of the set $\mathcal{A}^+$ was calculated and shown in Fig. 4.5(a) on a logarithmic scale. Similarly, the distribution of the optimal λ_T is illustrated in Fig. 4.5(b). As can be seen, λ_{GCV} is inclined to be closer to zero when compared to λ_T, while λ_T tends to be arranged more flatly. Also, to compare the absolute values of two optimal λ for the same voxel, the distribution of the difference $\Delta\lambda = \lambda_T - \lambda_{GCV}$ was calculated. The distribution of the difference value $\Delta\lambda$ plotted in Fig. 4.6 turned out to be skewed dramatically to the right, revealing that on average, the optimal value λ_T is larger than the optimal λ_{GCV}.

As temporal patterns of a typical hemodynamic response (Ogawa et al., 1998) are expected to be obtained, it is of crucial importance to investigate the effect of an

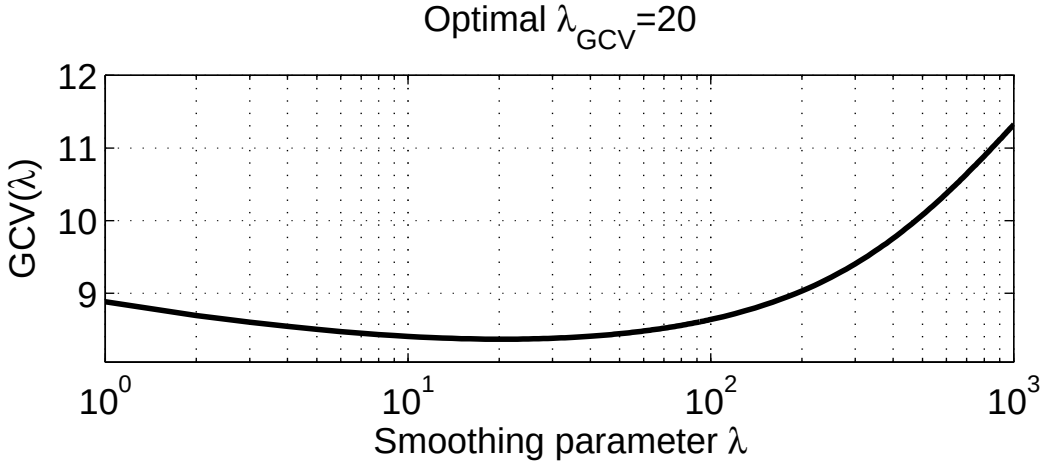


(a) Conventional, positive averaged hemodynamic responses.

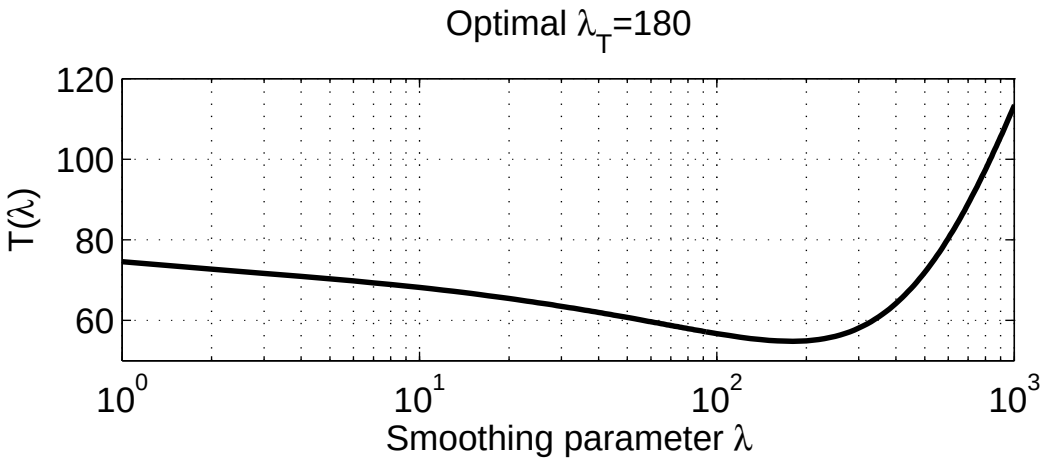


(b) Averaged hemodynamic responses with the negative pattern.

Figure 4.3: Illustrated are two activation patterns of the mean hemodynamic response for four voxels found to be activated in the ordinary BOLDfold analysis. The means for the fMRI signals were computed over repeated presentation blocks for each voxel separately. As can be seen, the responses may vary in shape and intensity. It should be noted that the two cases were easily discerned based on the value of the first derivative of the hemodynamic response at the initial point. The smoothing FDA techniques provide information on derivatives of the hemodynamic response at any point.

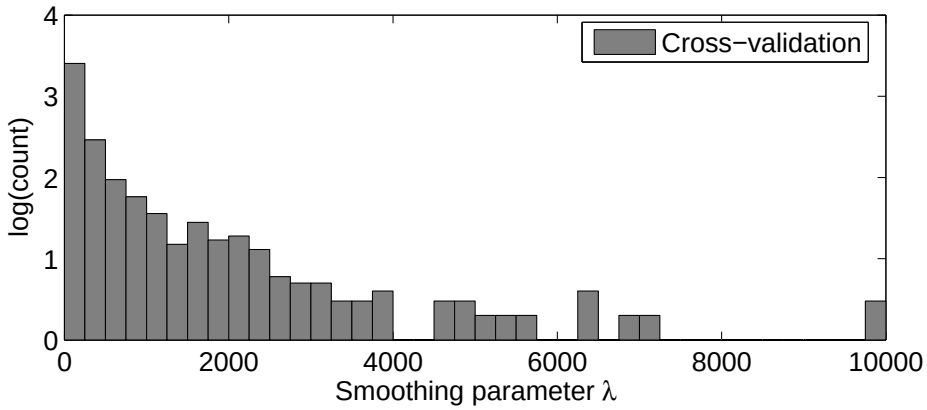


(a) The cross-validation score $GCV(\lambda)$

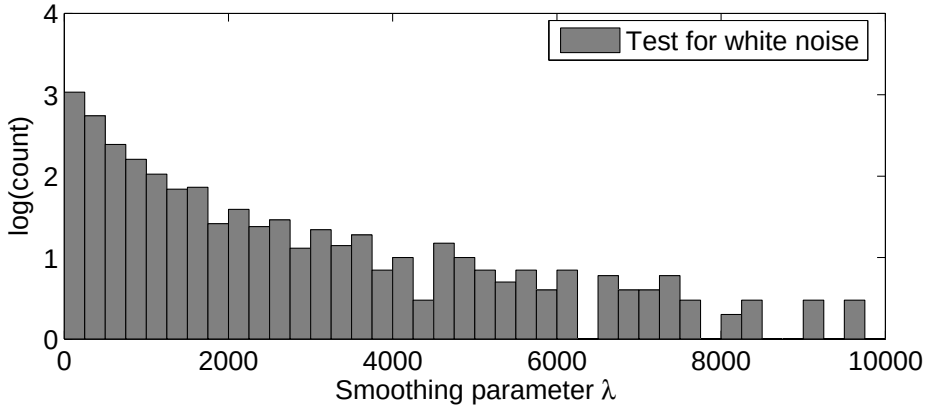


(b) The generalized χ^2 test value $T(\lambda)$.

Figure 4.4: The cross-validation score $GCV(\lambda)$ and the generalized χ^2 test value $T(\lambda)$ as functions of the regularization parameter λ for an activated voxel.



(a) The optimal λ_{GCV} distribution.



(b) The optimal λ_T distribution.

Figure 4.5: Logarithmic histograms showing the distributions of the optimal λ_{GCV} and λ_T. The activated voxels of the whole data set, demonstrating the conventional, positive response were employed to evaluate the λ distributions.

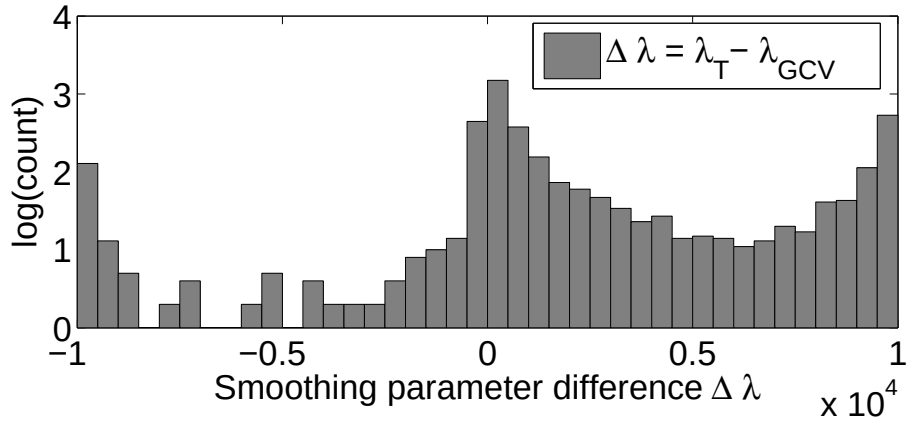
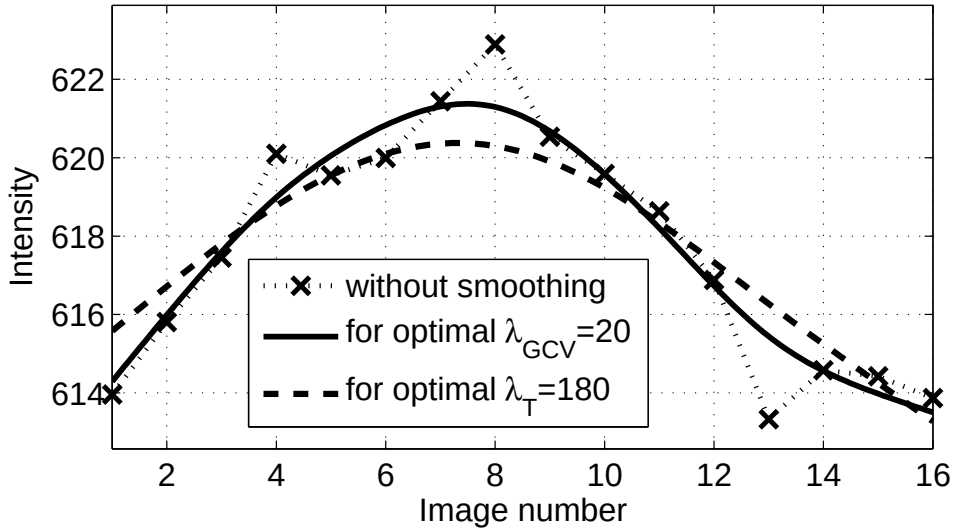


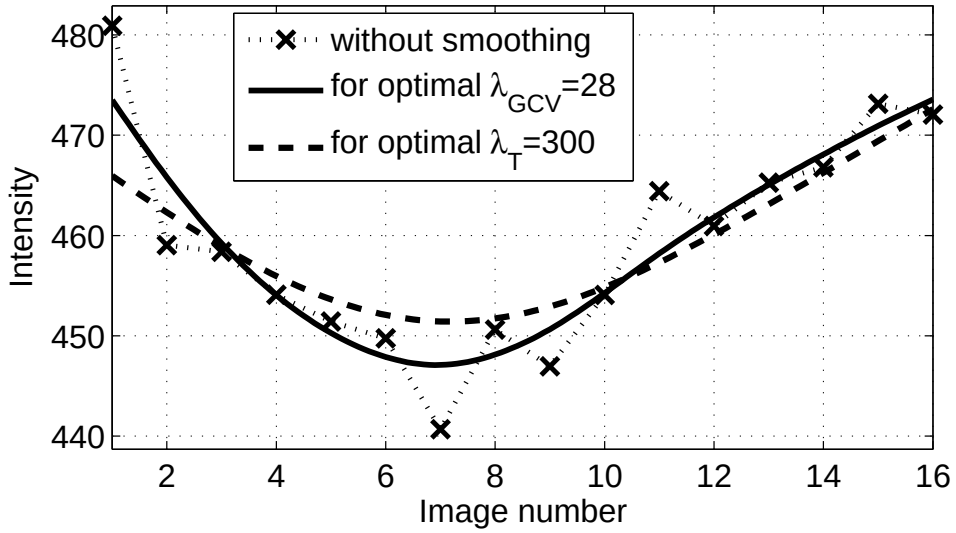
Figure 4.6: Logarithmic histogram showing the distribution of the difference $\Delta\lambda = \lambda_{\mathrm{T}} - \lambda_{GCV}$.

optimal λ on the averaged hemodynamic response. In an attempt to find a common basis for comparing the cross-validation and whitening criteria, we computed the smoothed hemodynamic responses for the optimal values λ_T and λ_{GCV}. Fig. 4.7(a) provides an example of the unsmoothed mean time course $\mathbf{h}$ of a voxel with a positive activation pattern, computed through the simple averaging (4.1) and two smoothed versions of the function $h(t)$. Similarly, an example of a negative mean fMRI response is illustrated in Fig. 4.7(b). The response corresponding to the smaller value of optimal λ_{GCV} in comparison to λ_T has larger curvature and more pronounced inflections than the function given by optimal λ_T.

What can be drawn from a visual analysis of the smoothed fMRI signal is that the expected shape of the hemodynamic response implicitly puts additional constraints on the definitional domain and optimality range of the value of λ. A typical hemodynamic response might display an initial dip, reach its maximum within a few seconds after the beginning of the stimulus and then decay after the termination of the stimulus with an undershoot in the time course (Ogawa et al., 1998). With the

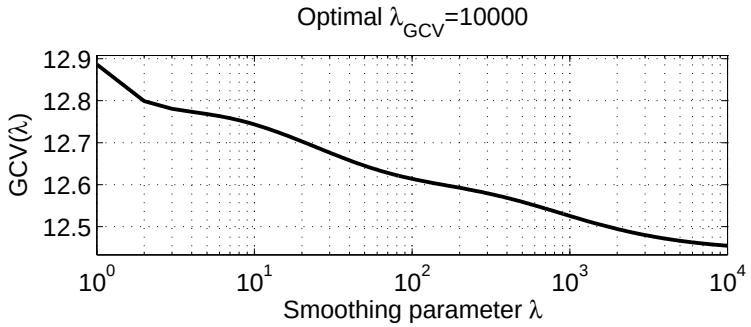


(a) An example of the positive activation pattern of the hemodynamic response. The functions $GCV(\lambda)$ and $T(\lambda)$ associated with this voxel are illustrated in Fig. 4.4

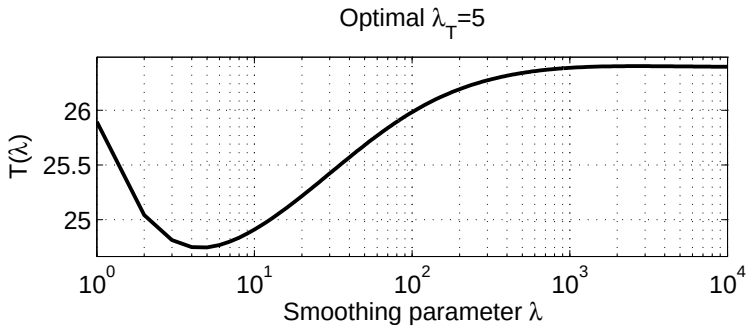


(b) An example of the negative activation pattern of the hemodynamic response.

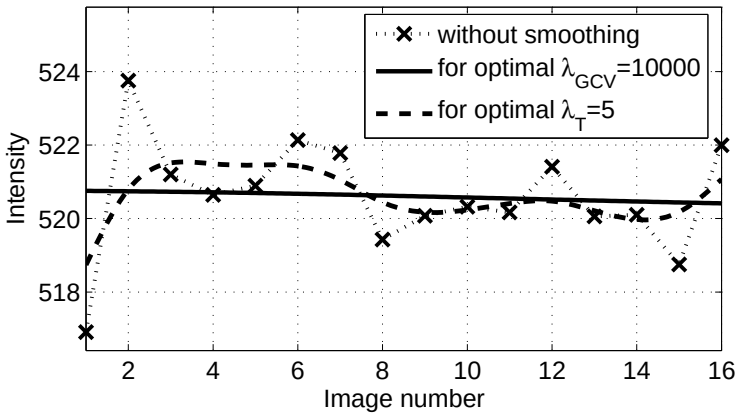
Figure 4.7: The averaged fMRI response for the same voxel, computed in three different ways. The noisy, discrete time series represents the mean response $\mathbf{h}$ calculated the simple averaging procedure (4.1). The solid line shows the smoothed function $h(t)$ evaluated through optimizing the generalized cross-validation value $GCV(\lambda)$, while the dashed line stands for the smoothed $h(t)$ obtained through minimizing the χ^2 value $T(\lambda)$.



(a) The generalized cross-validation score $GCV(\lambda)$.



(b) The generalized χ^2-test value $T(\lambda)$.



(c) The averaged hemodynamic response.

Figure 4.8: Illustrated is the relations between λ found to be optimal based on two criteria, GCV-score and the generalized χ^2-test for white noise, and their effect on the averaged hemodynamic response for one activated voxel. On average, GCV was found to supply less amount of smoothing then the χ^2-test. However, as can be seen, the GCV criterion can fail, while the alternative criterion seems to produce a reasonable amount of smoothing. Moreover, the averaged hemodynamic signal associated with the optimal λ_T demonstrates the typical activation features such as the rise, activation plateau and decay.

elusive nature of the initial dip (see Buxton (2001) for an excellent discussion), we did not expect to see the initial dip at a magnetic field of 1.5 T. Although a brief, weaker signal decrease before the main increase of the hemodynamic response was observed in some studies (for example, Yacoub et al. (2001)), we were not able to see an initial dip both in smoothed and original data. Nevertheless, the lower λ solutions more exactly match the expected response with rise, fall, undershoot and return to baseline. Functions $h(t)$ that correspond to high values of λ do not provide visually satisfactory fits because they have less inflection than expected. For this particular dataset, values of λ above approximately 200 introduce insufficient bending in the reconstructed hemodynamic response. It is characteristic that a large portion of all the values of λ optimized both through the cross-validation and whitening criteria were found to be in this range. It should be noted however that our judgement about the threshold for optimal λ is subjective. Furthermore, caution regarding expectations about the hemodynamic response should be taken as the fMRI signals have been shown to vary not only from subject to subject but also from voxel to voxel (Handwerker et al., 2004).

Finally, it is significant to note that there is no homogeneity in voxels regarding the amount of smoothing they experienced for both criteria. On one hand, there were voxels demonstrating that cross-validation works better than the criterion based on whitening the residuals from the perspective of the degree of inflection found in the smoothed hemodynamic response. On the other hand, the whitening criterion was sometimes able to produce a reasonable amount of smoothing (see Fig. 4.8(b)), while cross-validation failed because its minimum was found near very high values

of λ (see Fig. 4.8(a)) where the quadratic error term in (4.9) is ignored. To mitigate the driving force of cross-validation to increase the parameter λ, the data becomes oversmoothed up to being turned into a straight line (see Fig. 4.8).

4.5 Discussion

We applied functional data analysis methods to compute the smoothing of an averaged fMRI response $h(t)$ for repeated block design experiments. Smoothing techniques involving Tikhonov regularization were employed with the goal of removing the serially correlated observational errors in the GLM model. Specifically, the fMRI response $h(t)$ was expanded in cubic B-spline functions. The roughness of the hemodynamic response was quantified by the regularization term represented by the integrated curvature of the target function $h(t)$ with a proportionality coefficient λ. A tunable regularization parameter λ was used to regulate the trade-off between fit to the original time course, as measured by the residual sum of squares, and variability of the response $h(t)$, as determined by the regularization term. Furthermore, the parameter λ defined the amount of smoothing through the smoothing matrix $\mathbf{S}_\lambda$, which in turn controlled the autocorrelation structure of errors.

Two strategies were previously proposed in the literature to deal with serially correlated data, namely whitening the residuals and temporal filtering of the time series. It has been argued that whitening is a more efficient approach to parameter estimation, given that there is no discrepancy between the assumed and the actual autocorrelations (Friston et al., 2000). Since, in practice, the true autocorrelation

structure is never known, temporal filtering with the goal to smooth has been found to be a preferable technique (Friston et al., 2000).

In this work, we explored the possibility of combining the goals of the two strategies, namely to use the regularization parameter λ for whitening the residuals in a temporal smoothing scheme. This was possible due to the ability of the parameter λ to specify an optimal amount of smoothing for the noisy and discrete hemodynamic response. However, there is an ambiguity in determining what is considered optimal since the choice of the regularization parameter λ depends on the goal one wants to reach. We considered two criteria for the automatic selection of an optimal λ. The first criterion was the generalized cross-validation which compromises approximation and bending (Carew et al., 2003). It was compared to the approach for whitening residuals, based on a generalized χ^2 test of residuals for white noise. Our empirical results allow one to arrive at several conclusions. First, the autocorrelation structure of residuals can be improved with appropriate selection of the regularization parameter. Second, the generalized cross-validation does not guarantee the best correlation structure of the residuals as determined by the white noise test. Furthermore, cross-validation, on average, yields lower values of the parameter λ. It should be noted that both of the two criteria provide a wide range of values for the regularization parameter. However, high values of the smoothing parameter resulted in a insufficient bending of the modeled hemodynamic response and give a very smooth fit that ignores the data's features. It was previously reported that GCV tends to yield under-smoothing (Gu, 2002). Nevertheless, from the perspective of the expected hemodynamic shape, a criterion ensuring lower values of the regularization param-

eter, namely cross-validation, appears to be preferable. In spite of that, we were able to observe voxels that can not be forced into the Procrustean bed of any of the criteria tested. All this may lay a basis to suggest that using any criteria, individually and blindly, is not enough for constructing an effective automatic procedure for selecting the regularization parameter, and that additional information or a mixture of techniques would probably be required.

The shape of the fMRI response has been previously found to be of separate interest as its temporal characteristics were demonstrated to correlate with the behavioral measures of cognitive tasks (Menon et al., 1998; Liu et al., 2004). The FDA smoothing techniques allow temporal features of the hemodynamic response, such as the width or the onset time to be easily computed for possible correlation to task accuracy and reaction time. Also, these methods lay a basis for establishing a compromise between optimally denoising the discrete time course and improving the noise autocorrelation structure. In addition, the tested methods allow one to compute the derivatives of the hemodynamic response for comparing fMRI data to behavioral and physiological models. In particular and as used here, information on the derivatives can be used to robustly discern the positive and negative activation patterns of the fMRI time series.

Bibliography

R. Borowsky, W.J. Owen, and G.E. Sarty. The role of the left hemisphere in motor control of touch: a functional magnetic resonance imaging analysis. *Brain Cogn.*, 49(1):96–101, 2002.

E. Bullmore, M. Brammer, S.C. Williams, S. Rabe-Hesketh, N. Janot, A. David, J. Mellers, R. Howard, and P. Sham. Statistical methods of estimation and inference for functional MR image analysis. *Magn Reson Med*, 35(2):261–77, 1996a.

E. Bullmore, M. Brammer, S.C. Williams, S. Rabe-Hesketh, N. Janot, A. David, J. Mellers, R. Howard, and P. Sham. Statistical methods of estimation and inference for functional MR image analysis. *Magn Reson Med*, 35(2):261–77, 1996b.

R.B. Buxton. The elusive initial dip. *Neuroimage*, 13:953–8, 2001.

R.B. Buxton, E.C Wong, and L.R. Frank. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn Reson Med*, 39(6): 855–864, 1998.

J.D. Carew, G. Wahba, X. Xie, E.V. Nordheim, and M.E. Meyerand. Optimal spline smoothing of fMRI time series by generalized cross-validation. *Neuroimage*, 18(4): 950–61, 2003.

J.M. Carnicer and J.M. Pena. Totally positive bases for shape preserving curve design and optimality of B-splines. *Comput. Aided Geom. Design*, 11(6):633 – 654, 1994.

B. Chanmond. *Modeling and inverse problems in image analysis*. Springer, 2003.

S. Clare, M. Humberstone, J. Hykin, L.D. Blumhardt, R. Bowtell, and P. Morris. Detecting activations in event-related fMRI using analysis of variance. *Magn Reson Med*, 42(6):1117–22, December 1999.

P. Craven and G. Wahba. Smoothing noisy data with spline smoothing. *Numer Math*, (31):377–403, 1979.

N. Davies, C.M. Triggs, and P. Newbold. Significance level of the Box-Piece portmanteau statistics in finite samples. *Biometrica*, 64:517–22, 1977.

C. de Boor. *A practical guide to Splines*. Springer-Verlag, 1978.

E.A. DeYoe, P. Bandettini, J. Neitz, D. Miller, and P. Winans. Functional magnetic resonance imaging (fMRI) of the human brain. *J Neurosci Methods*, 54(2):171–87, 1994.

R.L. Eubank. *Spline smoothing and nonparametric approximation*. Marcel Dekker, 1988.

R.S.J Frackowiak, K.J. Friston, C.D. Frith, R.J. Dolan, and J.C. Mazziotta. *Human Brain Function*. Toronto: Academic Press, 1997.

K.J. Friston, A.P. Holmes, J.B. Poline, P.J. Grasby, S.C. Williams, R.S. Frackowiak, and R. Turner. Analysis of fMRI time-series revisited. *Neuroimage*, 2(1):173–81, 1995a.

K.J. Friston, A.P. Holmes, K.J. Worsley, J.B. Poline, C.D. Frith, and R.S.J. Frackowiak. Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp*, 2(3):189–210, 1995b.

K.J. Friston, P. Fletcher, O. Josephs, A. Holmes, M.D. Rugg, and R. Turner. Event-related fMRI: characterizing differential responses. *Neuroimage*, 7(1):30–40, January 1998.

K.J. Friston, O Josephs, E. Zarahn, A.P. Holmes, S. Rouquette, and J. Poline. To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. *Neuroimage*, 12(2):196–208, 2000.

C. Goutte, F.A. Nielsen, and L.K. Hansen. Modeling the haemodynamic response in fMRI using smooth FIR filters. *IEEE Trans Med Imaging*, 19(12):1188–201, 2000.

B.M. Graham and A. Adler. Objective selection of hyperparameter for EIT. *Physiol Meas*, 27(5):S65–79, 2006.

C. Gu. *Smoothing spline ANOVA models*. Springer, 2002.

D.A. Handwerker, J.M. Ollinger, and M. D'Esposito. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage*, 21(4):1639–51, 2004.

N. Harel, S.P. Lee, T. Nagaoka, D.S. Kim, and S.G. Kim. Origin of negative blood oxygenation level-dependent fMRI signals. *J Cereb Blood Flow Metab*, 22(8):908–17, 2002.

R Lavarello, F. Kamalabadi, and W.D.Jr O'Brien. A regularized inverse approach to ultrasonic pulse-echo imaging. *IEEE Trans Med Imaging*, 25(6):712–22, 2006.

H.L. Liu, W.T. Liao, S.Y. Fang, T.C. Chu, and L.H. Tan. Correlation between temporal response of fMRI and fast reaction time in a language task. *Magn Reson Imaging*, 22(4):451–5, 2004.

G.M. Ljung and G.E.P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

H.H. Lu, C.M. Chen, and I.H. Yang. Cross-reference weighted least square estimates for positron emission tomography. *IEEE Trans Med Imaging*, 17(1):1–8, 1998.

R.S. Menon, D.C. Luknowsky, and J.S. Gati. Mental chronometry using latency-resolved functional MRI. *Proc Natl Acad Sci U S A*, 95(18):10902–7, 1998.

S. Ogawa, R.S. Menon, S.G. Kim, and K. Ugurbil. On the characteristics of functional magnetic resonance imaging of the brain. *Annu Rev Biophys Biomol Struct*, 27:447–74, 1998.

J.O Ramsay and B.W. Silverman. *Applied Functional Data Analysis*. Springer, 2002.

G.E. Sarty and R. Borowsky. Functional MRI activation maps from empirically defined curve fitting. *Concepts in Magnetic Resonance Part B*, 24B(1):46–55, 2005.

A. Shmuel, M. Augath, A. Oeltermann, and N.K. Logothetis. Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. *Nat Neurosci*, 9(4):569–77, 2006.

S. Sourbron, R. Luypaert, P. Van Schuerbeek, M. Dujardin, and T. Stadnik. Choice of the regularization parameter for perfusion quantification with MRI. *Phys Med Biol*, 49(14):3307–24, 2004.

G. Wahba. *Spline models for observational data*. SIAM, 1990.

K.J. Worsley and K.J. Friston. Analysis of fMRI time-series revisited–again. *Neuroimage*, 2(3):173–81, 1995.

E. Yacoub, A. Shmuel, J. Pfeuffer, P.F Van De Moortele, G. Adriany, P Andersen, J.T. Vaughan, H. Merkle, K. Ugurbil, and X. Hu. Imaging brain function in humans at 7 Tesla. *Magn Reson Med*, 45(4):588–94, 2001.

E. Zarahn, G.K. Aguirre, and M. D'Esposito. Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *Neuroimage*, 5(3):179–97, 1997.

Chapter 5

The Relationship between Naming Reaction Time and Functional MRI Parameters in Broca's Area, and Evidence for an Independent-Dual-Route Model of Reading Behavior and Neurophysiology

5.1 Preliminaries

This chapter is based on the manuscript titled "The Relationship between Naming Reaction Time and Functional MRI Parameters in Broca's Area, and Evidence for an Independent-Dual-Route Model of Reading Behavior and Neurophysiology" by Cummine,J., Borowsky,R., Vakorin,V.A., Bird,J. and Sarty,G.E., submitted to *Brain*. For the present study, a complex of Matlab© procedures was written to extract temporal characteristics of fMRI time series. From the mathematical point of view, the implementation the Matlab© commands designed to compute maps of temporal characteristics of BOLD signals was based on the smoothing techniques considered in Chapter 4. The rationale for this work was to demonstrate that the functional data analysis (FDA) techniques developed in Chapter 4 can reliably and

stably be used to characterize the BOLD response it terms of a few parameters such as BOLD width, time-to-peak, initial slope or signal intensity. The objective is to apply the FDA techniques to an investigation in contemporary cognitive psychology in an effort to show that neuroimaging data contain information relevant to behavior. Specifically, in the context of a dual-route model of reading, we show that behavioral relationships normally measured with reaction time and accuracy may also be measured using BOLD characterization parameters extracted using our FDA technique.

The correlation between behavioral response time and functional parameters of BOLD responses measured in Broca's area during a word identification task was analyzed as a function of four stimulus types: regular words (*e.g., hint*), irregular words (*e.g., pint*), non-words (*e.g., bint*) and pseudohomophones (a type of non-word which sounds like a real word when pronounced, *e.g., pynt*). The results revealed that the BOLD response width is uniquely related to behaviorial reaction time in the task of pseudohomophone reading. In addition, a condition of independence between lexical and sub-lexical routes for converting speech to print, previously tested for predicting regular word accuracy given irregular word and non-word/pseudohomophone naming accuracy, was found to hold for BOLD width, BOLD time-to-peak and BOLD intensity as well as reaction time. These findings support previously reported behavioral evidence for the hypothesis of independent dual-route models of reading. Our ability to extract parameters that characterize the BOLD response and to find statistically significant correlations with behavioral parameters, including predictive power when a specific cognitive model (the dual-route reading model) was tested, demonstrates

the stability and robustness of the FDA characterization method across experimental subjects.

5.2 Introduction

One of the main applications of blood oxygen level dependent (BOLD) fMRI in studying brain function is to trace the topography of activation in reaction to performing a cognitive, perceptual or motor task (see, for example, Frackowiak et al. (1997)). Typically, the mapping of brain activation relies on identifying stimulus-related changes in MRI signal intensity (Friston et al., 1994, 1995). However, such an approach implicitly assumes the static nature of the underlying activation processes. A next step towards a more advanced analysis is to study peculiarities of the BOLD dynamics *per se*. In that context, considered in Chapter 3, there is a growing interest in exploring the question of how functional interactions between brain regions evolve during task performance (Friston et al., 2003; Lee et al., 2006). Another possible approach, considered in this chapter, is to relate temporal characteristics of hemodynamic responses to behaviorial measures (Richter et al., 1997b; Menon and Kim, 1999). Under this paradigm, the question is how much behavioral information can be extracted from the amplitude and timing of hemodynamic responses. The amplitude and the timing of the hemodynamic response in turn can be difficult to determine because of the noisy nature of the measured signal. Thus we aim to extract reliable values for the amplitude and timing parameters using our previously developed FDA approach. A separate problem, not considered in this chapter, is to

separate the temporal characteristics of BOLD signals, resulting from differences in hemodynamic response functions from those caused by actual differences in neural activity (Aguirre et al. (1998); Handwerker et al. (2004)).

Previous reseachers have used a variety of approaches for characterizing the amplitude and timing of the BOLD response and correlating the neuroimaging data to behavioral data. Menon et al. (1998) collected rapid fMRI images in a cued visuomotor reaction time task to trace the sequence of brain activation, leading from $V1$ to the supplementary motor area (SMA) and to the primary motor area (M1). The activation pathway leading from visual area V1 to SMA was characterized by a constant delay in activation, irrespective of latency in reacting to visual stimuli. In contrast, the SMA-to-M1 delay was found to be linearly proportional to reaction time.

Using time-resolved fMRI, Richter et al. (1997a, 2000) investigated temporal sequences of neural events in motor and premotor areas during a mental rotation task. One of the questions considered in these studies probed the relationship between behaviorial response times and several functional parameters (onset latency and width) of the BOLD response in selected brain regions. The authors found the width of the fMRI response in the parietal lobe to be proportional to the reaction time (Richter et al., 1997a). The study by Richter et al. (2000) revealed a positive correlation between response time and the BOLD width in all the brain regions in question, except for the left M1.

Following Richter et al. (2000), Liu et al. (2004) sought to relate reaction time and onset time to the full-width-at-half-maximum (FWHM) of hemodynamic responses

in a task of more rapid cognitive demands (a lexical decision task to decide if the target is a word or a non-word). The study employed three type of stimuli: high-frequency words, low-frequency words and non-words. Liu et al. (2004) reported a significant positive correlation between the reaction time and the FWHM (but not for the onset time).

For the test application of our FDA technique for extracting BOLD parameters to correlate with behaviorial data we examined a dual-route model for reading. We now give a brief overview of this model. It has been suggested that two different mechanisms are available for skilled readers to convert speech to sound: via the lexical and the sub-lexical routes (Coltheart et al., 2001; Coltheart, 2006). Familiar regular words have lexical representations, as well as obeying typical spelling-sound rules. Due to such a double nature, both the lexical system and the sub-lexical system may independently perform the correct naming of regular words. A study by Castles et al. (2006) demonstrated that these two processes interact in a independent way. To support the independence of the lexical and sub-lexical systems, the authors proposed to use a normalized accuracy measure (ranging from 0 to 1) of reading regular words (REG), irregular words (IRR) and non-words (NW). Specifically, they showed that the actual REG accuracy A_{REG}^{act} can be predicted from the actual IRR accuracy A_{IRR} and the actual NW accuracy A_{NW} by the relation:

$$A_{\mathrm{REG}}^{pred} = A_{\mathrm{IRR}} + A_{\mathrm{NW}} - A_{\mathrm{IRR}} A_{\mathrm{NW}} \tag{5.1}$$

where A_{REG}^{pred} is the predicted REG accuracy.

Given the discrete and noisy nature of the fMRI response, extracting temporal pa-

rameters of the hemodynamic response inexorably requires some kind of smoothing. As noted in Chapter 4, the ultimate goal of the proposed techniques was to verify and extend previously proposed correlations between behavioral and functional measures. This study illustrates the usefulness of the B-splines-based smoothing techniques considered in Chapter 4, in refining our understanding of the relationships between word identification and neurophysiological processes in the left inferior frontal gyrus (namely, Broca's area).

From the cognitive science point-of-view, it would be of interest to determine if the independence of the lexical and sub-lexical systems, stated in terms of accuracy, can be extended to temporal and spatial parameters that characterize the hemodynamic response. The list of possible candidates includes FWHM, time-to-peak, BOLD signal intensity, volume (expressed in number of activated voxels, see Halari et al. (2006)), parameters related to the derivatives of BOLD response, such as initial slope (the first derivative at the initial point) and inflection point (a time point at which the second derivative is equal to zero, see Menon et al. (1998)). Also, it would be of interest to refine the findings of Liu et al. (2004) by exploring the question of whether the nature of correlations between reaction time and FWHM of the BOLD response in Broca's area depends on stimulus type, including the processing of pseudohomophones (or non-words) and irregular words in a naming task. Our main results based on using fMRI data smoothed with B-splines as described in Chapter 4, are reviewed in Section 5.6.

5.3 Independence of lexical and sub-lexical systems

We will first define a set of the variables of interest. The hemodynamic response $h(t)$, where $t \in [t_1, t_T]$, averaged over blocks in a block designed experiment can be parameterized in terms of basis functions $\phi_j(t)$ as follows

$$h(t) = \sum_{j=1}^{K} c_j \, \phi_j \, (t) \tag{5.2}$$

where $\{\phi_j\}_{j=1}^{K}$ is a set of B-spline basis functions (see Section 4.3.1). Let the BOLD response $h(t)$ at the initial point t_1 be equal to h_1. Also, let the hemodynamic signal $h(t)$ reach its maximum value $h_{max} = \max_t h(t)$ at the point $t_{max} = \arg\max_t h(t)$.

BOLD Width (BW) or Full-Width-at-Half-Maximum (FWHM) is the distance between points t_L and t_R, $t_L < t_R$, on the hemodynamic response curve h(t) at which the intensity of the hemodynamic response reaches half its maximum value. Specifically, if $t_{L,R} = \arg\min_t |\frac{h_{max}-h_1}{2} - h(t)|$, then BW $= t_R - t_L$. A note here is that the resolution of BW is not limited by T_R. Time-to-Peak, Δ, is the period from the initial point to the time point where the hemodynamic response reaches its maximum value, specifically $\Delta = t_{max} - t_1$. Intensity, I, is defined as the maximum value of the hemodynamic signal, namely I $= h_{max}$. Initial slope, S, is the value of the first derivative of the hemodynamic response at the initial point t_1, specifically $\mathrm{S} = \frac{\mathrm{d}h(t)}{\mathrm{d}t}|_{t=t_1} = \sum_{j=1}^{K} c_j \, \phi_j' \, (t_1)$. Finally, volume, V, is defined in terms of the number of activated voxels. The volume occupied by one voxel is equal to $2\,\mathrm{mm} \times 2\,\mathrm{mm} \times 8\,\mathrm{mm} = 32\,\mathrm{mm}^3$ (see Section 5.4).

The condition of independence between the lexical and sub-lexical systems, tested by Castles et al. (2006), connotes that prior to the main analysis, tested parameters should be normalized to range from 0 and 1. To convert the values of the BOLD FWHM (BW) variable to the interval $[0, 1]$, the actual BWs were normalized relative to the period interval 70.3 s. Let BW_{REG}, BW_{IRR}, BW_{NW} and BW_{PH} denote the BOLD FWHM for regular words, irregular words, non-words and pseudohomophones, respectively. Also, let the upper indices "act" and "pred" designate actual and predicted values, respectively. Similar to the accuracy-based independence model in (5.1), regular word BW can be predicted by

$$BW_{\mathrm{REG}}^{pred} = BW_{\mathrm{IRR}} + BW_{\mathrm{NW}} - BW_{\mathrm{IRR}} BW_{\mathrm{NW}} \tag{5.3}$$

based on non-words, and

$$BW_{\mathrm{REG}}^{pred} = BW_{\mathrm{IRR}} + BW_{\mathrm{PH}} - BW_{\mathrm{IRR}} BW_{\mathrm{PH}} \tag{5.4}$$

based on pseudohomophones.

The degree to which BW_{REG}^{act} can be explained by BW_{REG}^{pred}, is assessed through a linear regression without a constant:

$$BW_{\mathrm{REG}}^{act} = \beta_{BW}\ BW_{\mathrm{REG}}^{pred} + \varepsilon_W \tag{5.5}$$

where the proportionality coefficient β_{BW} represents the slope, and $\boldsymbol{\varepsilon}$ is zero mean, uncorrelated random error with common variance between subjects.

Similar to BOLD FWHM, time-to-peak, intensity, volume and initial slope can be used to test the model of independence between the lexical and sub-lexical systems. Before further analysis, these variables were normalized to their maximum possible

values. Specifically, the values of the parameters FWHM and time-to-peak were divided by the time of one period (19 volumes $\times\, T_R = 70.3$ s) similar to FWHM. In order to normalize intensity, volume and initial slope, the maximum obtained values were identified to be 1000 greyscale units for signal intensity, 150 cm^3 for volume and 10 grayscale units per s for initial slope. Similar to the model (5.5) explaining the actual BOLD width from the predicted, regression analyses were conducted for the variables time-to-peak Δ, intensity I, volume V and initial slope S:

$$\Delta_{\mathrm{REG}}^{act} = \beta_{\Delta}\ \Delta_{\mathrm{REG}}^{pred} + \varepsilon_{\Delta} \tag{5.6}$$

$$I_{\mathrm{REG}}^{act} = \beta_{I}\ I_{\mathrm{REG}}^{pred} + \varepsilon_{I} \tag{5.7}$$

$$V_{\mathrm{REG}}^{act} = \beta_{V}\ V_{\mathrm{REG}}^{pred} + \varepsilon_{V} \tag{5.8}$$

$$S_{\mathrm{REG}}^{act} = \beta_{S}\ S_{\mathrm{REG}}^{pred} + \varepsilon_{S} \tag{5.9}$$

where $\boldsymbol{\varepsilon}.$ is zero mean, uncorrelated random error with common variance between subjects.

5.4 fMRI experiments

Ten University graduate students (mean age of 25 years) participated in this study. There were four categories of letter strings presented: 55 for each category (regular

words, irregular words, non-words and pseudohomophones) for a total of 220 letter strings. Stimuli were matched, based on onset phoneme, length and word frequency for REG, IRR, and PH stimuli (for more details see McDougall et al. (2005)). Image acquisition was synchronized with visually presented stimuli and triggered by EPrime software (Psychology Software Tools, Inc., Pittsburgh, USA). The stimuli were projected with an LCD projector interfaced with the computer running the EPrime software. A mirror was attached to the MRI head coil to make a back-projection screen visible to the experimental participant. To minimize movement-related artifacts, responses were collected over the MRI intercom during the gaps in image acquisition. The experimenter used a button press to identify the moment when the participant completed the naming response.

A gradient echo T_2^* single-shot EPI sequence with fat saturation was used for BOLD measurement. Depending on distance between the posterior commissure and the top of the brain for each participant, either the 3^{rd} or 4^{th} inferior-most slice was centered on the posterior commissure in order to cover the entire cortex with one volume consisting of 12 slices. The imaging parameters were as follows: a 250 mm square field of view, 8 mm slice thickness, 2 mm interslice separation, $T_E = 55$ms and $T_R = 3700$ms with a 1650 ms gap of no acquisition in each repetition ($\frac{1}{2}$ of T_R during which the overt responses were made). For anatomical detail, T_1-weighted high-resolution spin warp spin-echo anatomical images ($T_R = 400$ ms, $T_E = 12$ ms, 256 x 256 acquisition matrix) were acquired in axial, sagittal, and coronal orientations for each subject with a slice thickness of 8 mm, interslice distance of 2 mm with the location, of axial images, being identical to that used in the functional imaging.

The four naming tasks were conducted under the blocked design paradigm. One block contained 19 volumes: the first 8 volumes were stimuli, followed by 11 volumes of rest. The block pattern was repeated 5 times. The first 5 image volumes, collected at the beginning to allow the spins to reach a steady state, were discarded prior to analysis. The complete data set consisted of 100 image volumes of which 95 were used in the analysis. Voxel time series with an intensity below a cut-off of 200 grey-scale units, being considered background, were excluded from further analysis.

5.5 Analysis

For optimal sensitivity, the experiment used a blocked design, as described above, and was analyzed using a previously developed and validated regularized BOLDfold approach (Borowsky et al., 2002, 2005; Sarty and Borowsky, 2005; Vakorin et al., 2006). The method adopted here is within the framework of signal averaging techniques that increase the signal-to-noise ratio of fMRI signals (DeYoe et al., 1994). In order to obtain a smoothed fMRI signal for a block, we incorporated spline interpolation in our analysis method. With the spline functions incorporated into a general linear model (GLM) approach, we effectively computed the hemodynamic response averaged over all the trials and used that as a reference function for computing the activation maps (Vakorin et al. (2006), Chapter 4 this thesis). The method is a generalization of the BOLDfold approach, which is suitable for repeated block design experiments in which the comparison function, assumed to be different for every voxel, is defined empirically through periodic data folding.

In BOLDfold, which was applied as our first analysis step, the voxelwise-calculated activation maps were based on correlations between the intensity time course and its repeated mean time course. Specifically, let y denote the observed time series with the number of sample points equal to $P \times T$ where P (equal to 5 here) represents the number of repeated imaging blocks (periods), and T (equal to 19 here) is the number of time point samples (equal to number of volumes) for each block. Also, let h denote the averaged, folded data (T points) over P periods. The averaged signal h can be repeated P times producing a reference signal $\tilde{h}$ spanned over the whole experiment. This newly constructed signal $\tilde{h}$ was correlated with the original time series y. High correlation r between y and $\tilde{h}$ reflects the activation of a specific brain area in response to the given stimulus. To define activated regions, a threshold correlation of 0.65 was used, with voxels having $r > 0.65$ being declared as active.

Next, to extract the relevant BOLD data, regions of interest (ROIs) were delineated for each participant (see Fig. 5.1). The ROIs were in the lateral inferior frontal gyrus of both the right and left hemisphere. For activated voxels in the delineated ROIs, the hemodynamic responses, h, although averaged now over P blocks but still discrete and noisy data, were fitted to a continuous function $h(t)$ expanded as a linear combination of B-spline basis functions (de Boor, 1978), acting as a filter on the noisy data. The coefficients of the linear expansion of $h(t)$ were estimated within the framework of a GLM with a criterion to minimize the total quadratic errors between the original time series y and the modeled function $h(t)$ unfolded P times. In addition, the GLM model (and the original BOLDfold model) were corrected for non-specific linear signal drift. As shown in Chapter 4, caution should

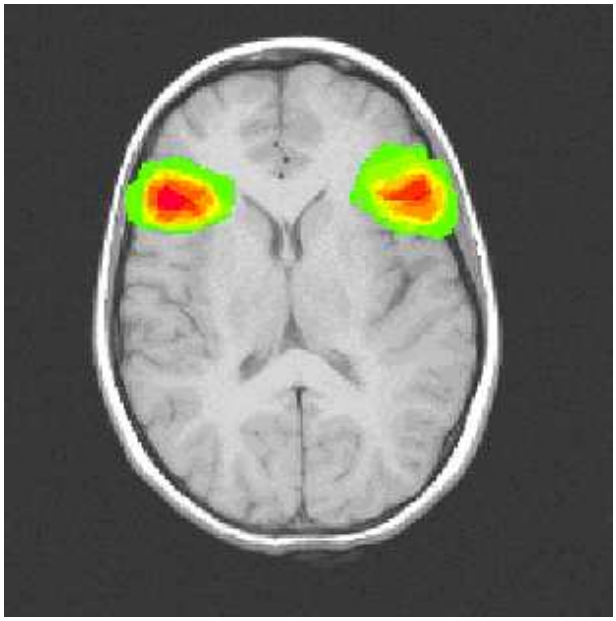


Figure 5.1: Inferior frontal gyri (Broca's Area), superimposed on the axial anatomical image of one of the participants. The amount of individual variability are portrayed with different grey-scale gradations. The regions masked in red represent common activation area for all the participants.

be taken in choosing blindly the optimal value of the regularization parameter since there is no guarantee against very large values of the regularization parameter, and hence, very strong smoothing, which eliminates temporal features of the hemodynamic response. We can avoid this problem by using $\lambda = 0$ and with the number of basis functions K set to be less than the number of sample points T. A reasonable default is to choose K to ensure that $T/K = 4 \ldots 5$ with the knots placed uniformly on the interval (Ruppert et al., 2003). For the present study we used $K = 6$. The smoothed hemodynamic responses, $h(t)$, calculated with high temporal resolution, were used to create the FWHM maps as well as maps for the first derivative at the initial point (i.e., initial slope of the BOLD response), intensity (maximum response value), and time-to-peak. Next, these four hemodynamic response parameter values

were averaged over all the active voxels included in the ROIs. It is these ROI averaged hemodynamic response parameters that were used for further analysis of the relationships between reaction time and the properties of the hemodynamic response.

A note here is that active regions were defined through regular BOLDfold analysis rather than using the smoothed BOLDfold method when the smoothed averaged function $h(t)$ unfolded P times correlates with the original time series y. In fact, this issue relates to the trade-off between detection power, or knowing what voxel are activated, and estimation efficiency, or knowing the time course of an activated voxel (Liu, 2004; Liu and Frank, 2004). As shown in Chapter 4, using either cross-validation or a white noise test does not protect the hemodynamic responses of activated voxels against oversmoothing. So oversmoothing may be a problem for detection. Nevertheless, it is not a significant problem for BOLD characterization.

5.6 Review of the main results

5.6.1 Relationship between Reaction Time and FWHM

Table 5.1 summarizes the mean values of reaction time and several BOLD response parameters for each stimulus type. Averaging the BOLD parameters across stimulus type, we found a significant correlation between Reaction Time and Volume ($r(39) = 0.32, p = 0.042$), Reaction Time and Initial Slope ($r(39) = 0.46, p = 0.003$). Although small, the correlation between Reaction Time and FWHM was found to be statistically significant at the 95% level, namely $r(39) = 0.269, p = 0.047$ (one-tailed

Parameters	Regular Words	Irregular Words	Non-words	Pseudo-homophones
FWHM, s	49.32	46.53	52.61	56.50
Intensity, *greyscale units*	666.70	702.36	716.94	710.92
Time-to-peak, s	41.99	41.63	42.81	43.96
Initial Slope, *greyscale units per s*	3.07	2.63	4.31	3.84
Volume, cm^3	36.00	35.3	70.50	66.80
Reaction time, ms	794.75	938.55	1084.30	1008.40

Table 5.1: Reaction time and the mean parameters of BOLD responses, as a function of stimulus type.

test), which is consistent with Liu et al. (2004).

In addition, we conducted a correlation analysis to study relationships between Reaction Time and BOLD parameters separately for each stimulus type: REG, IRR, NW and PH. Reaction Time for reading aloud pseudohomophones was found to be correlated with FWHM ($r(39) = 0.90, p < 0.001$), Intensity ($r(39) = 0.74, p < 0.015$) and Initial Slope ($r(39) = 0.69, p < 0.027$). A correlation analysis revealed no significant relationship between Reaction Time and the BOLD parameters for REG, IRR and NW, except for NW Initial Slope ($r(39) = -0.66, p < 0.036$).

On the whole, the aforementioned results supported the hypothesis that PH reaction time is correlated to FWHM, Intensity, Time-to-Peak, Volume and Initial Slope of the BOLD responses. The overall multiple regression of PH reaction time on five considered BOLD parameters was found to be significant; $F(4, 9) = 6.95$, $p = 0.04$.

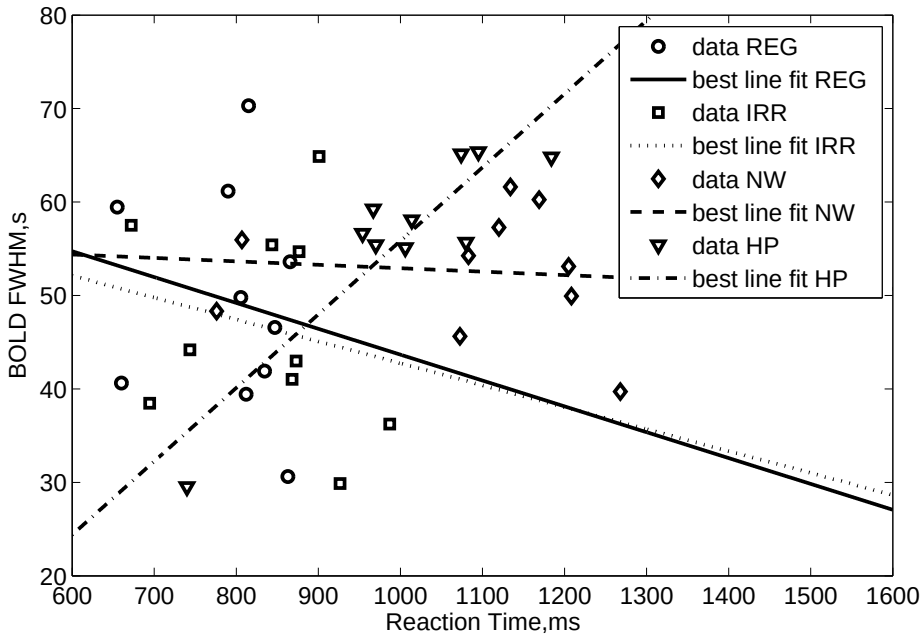


Figure 5.2: The Full-Width-at-Half-Maximum(FWHM) of BOLD responses as a function of reaction time for each stimulus type: regular words (REG), irregular words (IRR), non-words (NW) and pseudohomophones (PH). There is a significant positive correlation between the BOLD FWHM and reaction time in a task for overtly reading pseudohomophones. In contrast to PH, other types of stimuli did not show such a relationship.

However, only the FWHM-related coefficient in the overall model accounted for significantly unique variance: $t(9) = 3.21, p < 0.033$. Fig. 5.2 illustrates the significant correlation between the BOLD FWHM and Reaction Time for reading aloud pseudohomophones. Other types of stimuli failed to demonstrate such a relationship (Fig. 5.2).

5.6.2 BOLD parameters and the dual route reading model

The linear model was used to assess degree to which predicted Regular Word BOLD Full-Width-at-Half-Maximum BW^{pred}_{REG} can explain actual Regular Word BOLD FWHM BW^{act}_{REG}. For the independence model with pseudohomophones, it was estimated that the coefficient of determination R^2, which is equal, in this simple model, to the squared value r^2 of the Pearson correlation between BW^{act}_{REG} and BW^{pred}_{REG}, is 0.38 (R^2 adjusted for the number of explanatory terms in a model, R^2_{adj}, is equal to 0.30). For the independence model with NWs, $R^2 = .67$ ($R^2_{adj} = 0.63$). The slope coefficient was found to be statistically significant: at the 90% confidence level ($\alpha = 0.10$) for pseudohomophones with $t(9) = 2.20$ and $p = 0.056$, and with the 95% confidence ($\alpha = 0.05$) for non-words with $t(9) = 4.02$ and $p = 0.004$ (Table 5.2).

Also, a linear regression model was tested to see how well actual Regular Word BOLD Intensity, I^{act}_{REG}, can be accounted for with predicted Word BOLD Intensity I^{pred}_{REG}. We found that using both pseudohomophones and non-words in the independence model, the squared valued of the correlation between I^{pred}_{REG} and I^{act}_{REG} is about 0.50 (with $R^2_{adj} = 0.44$). Furthermore, not much difference between non-words and pseudohomophones was found in estimating the slope coefficient. Specifically, the

Model	R^2_{adj}	β_{BW} (Slope)	t-statistics	d.f.	p-value
BW^{act}_{REG} [IRR & PH]	0.30	1.38	2.21	9	0.058
BW^{act}_{REG} [IRR & NW]	0.63	2.15	4.02	9	0.004

Table 5.2: Regression of Actual Regular Word BOLD Full-Width-at-Half-Maximum $BW_{\text{REG}^{act}}$ on Predicted Regular BOLD Full-Width-at-Half-Maximum (FWHM) $BW_{\text{REG}^{pred}}$. The model of independence between lexical and sub-lexical routes was tested using FWHM.

Model	Adj. R^2	Slope	t-statistics	d.f.	p-value
I^{act}_{REG} [IRR & PH]	0.44	0.90	2.83	9	0.022
I^{act}_{REG} [IRR & NW]	0.44	1.41	2.85	9	0.021

Table 5.3: Regression of Actual Regular Word BOLD Intensity I^{act}_{REG} on Predicted Regular BOLD Intensity I^{pred}_{REG}, which tested the Intensity-based model of independence between lexical and sub-lexical systems.

slope was significant in both cases: $t(9) = 2.83$ and $p = 0.022$ for pseudohomophones, and $t(9) = 2.85$ and $p = 0.021$ for non-words (see Table 5.3).

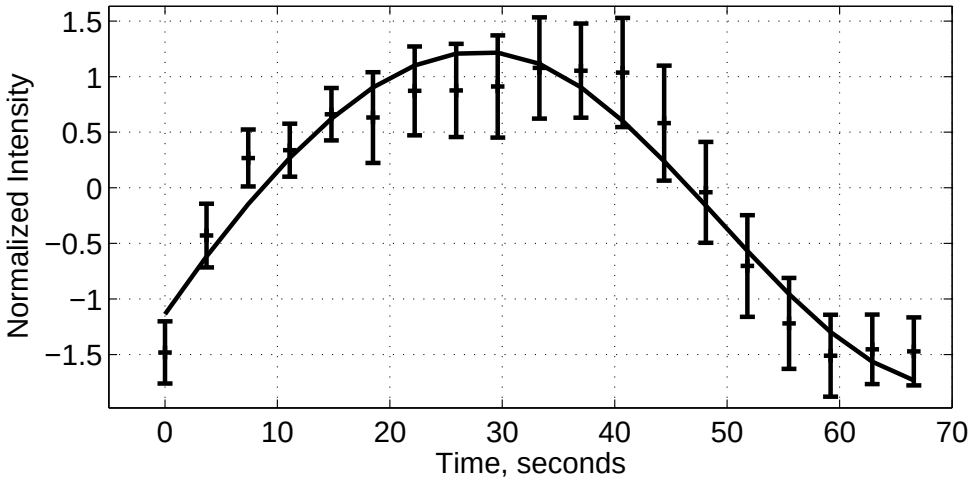
Further, a linear regression analysis was carried out to test the amount of variance in actual Regular World BOLD Time-to-Peak $\Delta^{act}_{\text{REG}}$ explained by the variable predicted Regular World BOLD $\Delta^{pred}_{\text{REG}}$. We found that $R^2_{adj} = 0.47$, $t(9) = 2.99$ and $p = .017$ for pseudohomophones, $R^2_{adj} = 0.38$, $t(9) = 2.53$ and $p = .035$ for non-words (see Table 5.4). In both cases, the slope coefficient was significant at the 95% confidence level.

Finally, we assessed the degree to which actual volume and initial slope can be accounted for by the predicted volume and initial slope, respectively. We found no significant correlation between actual volume and predicted volume for both pseudo-

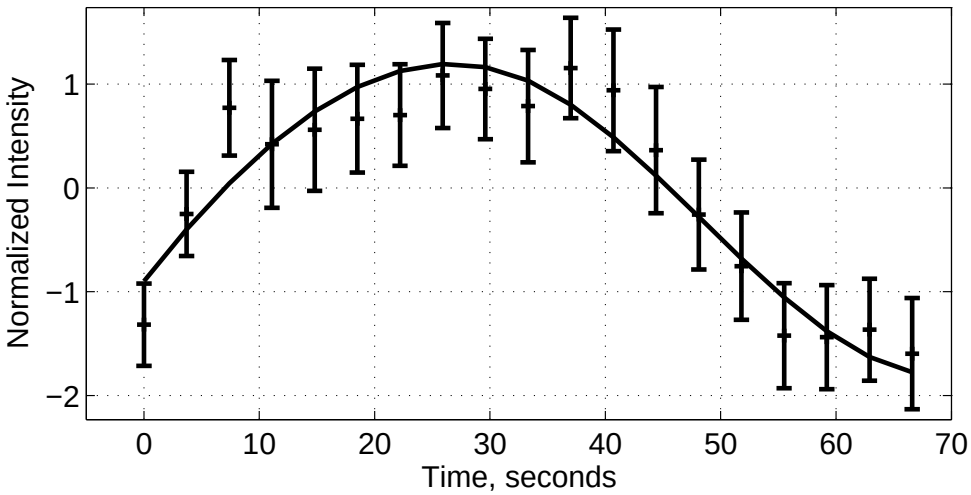
Model	Adj. R^2	Slope	t-statistics	d.f.	p-value
$\Delta^{act}_{\text{REG}}$ [IRR & PH]	0.47	1.65	2.99	9	0.017
$\Delta^{act}_{\text{REG}}$ [IRR & NW]	0.38	1.14	2.53	9	0.035

Table 5.4: Regression of Actual Regular Word BOLD Time-to-Peak $\Delta^{act}_{\text{REG}}$ on Predicted Regular BOLD Time-to-Peak $\Delta^{pred}_{\text{REG}}$, which tested the Time-to-peak-based model of independence between lexical and sub-lexical mechanisms of reading words.

homophones ($t(9) = 1.57$ and $p = 0.16$) and non-words ($t(9) = 0.75$ and $p = 0.48$). Similar to volume, the independence model based on initial slope did not reveal any significant relationship for both pseudohomophones ($t(9) = 1.65$ and $p = 0.14$) and non-words ($t(9) = 1.18$ and $p = 0.27$).

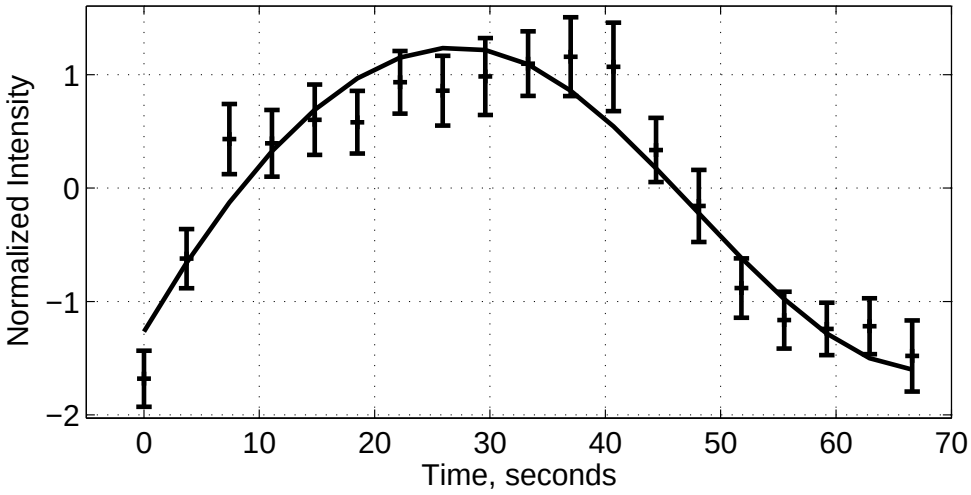


(a) Regular words

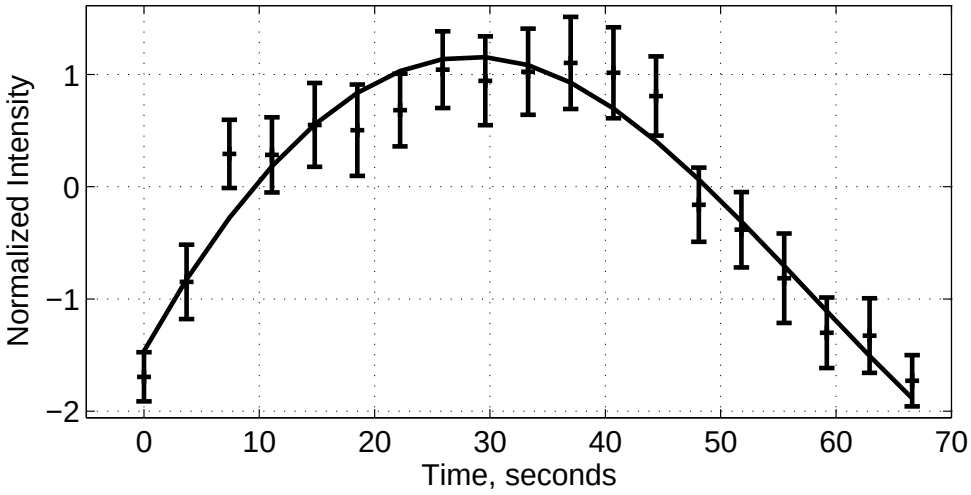


(b) Irregular words

Figure 5.3: Observed and smoothed hemodynamic responses in Broca's area in reaction to reading regular and irregular words. The hemodynamic responses are smoothed across participants. Error bars indicate the range of one standard deviation. The BOLD signal intensity is shown in terms of standardized (as z-scores) grayscale intensity values.



(a) Non-words



(b) Pseudohomophones

Figure 5.4: Observed and smoothed hemodynamic responses in Broca's area in reaction to reading non-words and pseudohomophones. The hemodynamic responses are smoothed across participants. Error bars indicate the range of one standard deviation. The BOLD signal intensity is shown in terms of standardized (as z-scores) grayscale intensity values.

5.7 Discussion and conclusion

There are two behavioral results from this study that we would like to point out. First, there exists a linear relationship between reaction time and FWHM of the hemodynamic response in a naming task, and this relationship depends on stimulus type. Specifically, only reaction time in a task of reading pseudohomophones was found to be significantly related to the BOLD width. Second, the findings support previously reported behavioral evidence for an independent dual-route model of reading. The current study has demonstrated that the condition for mathematical independence between the two pathways for converting print to speech holds for a set of temporal characteristics of BOLD signals, namely, the signal width, signal intensity and time-to-peak. Specifically, regular word response time, regular word BOLD FWHM, regular word time-to-peak and regular word BOLD intensity can be predicted through the corresponding irregular word and non-word or pseudohomophone parameters.

The basis for the behavioral findings revealed in this chapter is the smoothing techniques introduced in Chapter 4. Figures 5.3 and 5.4 represent hemodynamic responses in reaction to reading regular words, irregular words, non-words and pseudohomophones as found in Broca's area and averaged across participants. As one can see, the presence of noise in the averaged BOLD responses necessitates use of smoothing techniques before computing maps of BOLD parameters in question.

Strictly speaking, the smoothing techniques used in this chapter are a modification of the methods used in Chapter 4 with the regularization parameter λ set to

be equal zero. The main reason for this is an unresolved issue on what criterion for selecting the optimal regularization parameter is best suited for determining the "right" amount of smoothing. The set of possible candidate may include a white noise test, cross-validation (or generalized cross-validation), Akaike information criterion, Mallow's C_p-based criterion. All these principles do not ensure that the hemodynamic responses of activated voxels cannot be oversmoothed. To circumvent this issue, we can artificially set $\lambda = 0$, which, in turn, requires that the number of basis functions K must be less than the number of sample points T in order to avoid the multicollinearity (see Chapter 4, paragraph 4.3.1). As indicated in Ruppert et al. (2003), the ratio T/K equal to $4 \ldots 5$ would be a reasonable default. Having $\lambda = 0$ does not guarantee the best residual auto-correlation structure as determined by a χ^2 test of residuals for white noise, which results in a bias between the true and assumed autocorrelation. Nevertheless, smoothing as an alternative to an autoregressive AR(1) model (Bullmore et al., 1996) or a modified $1/f$ model (Zarahn et al., 1997), provides an acceptable level of bias (Friston et al., 2000). Having some autocorrelation in the noise term appears to be a lesser problem than getting a stable smooth estimate of the BOLD response shape.

The results show that the smoothing techniques for extracting parameters, like BOLD width, that characterize the BOLD response are stable enough to support correlations normally seen only in behavioral data. In particular, a mathematical characterization of a cognitive model of reading which is normally based on reaction time was shown to hold also for BOLD width, initial slope and time-to-peak. This work opens the possibility that BOLD-based neuroimaging may be used for cognitive

science in a more direct way to test hypotheses about cognition than it has been used in the past.

Bibliography

G.K. Aguirre, E. Zarahn, and M. D'esposito. The variability of human BOLD hemodynamic responses. *Neuroimage*, 8(4):360–369, 1998.

R. Borowsky, W.J. Owen, and G.E. Sarty. The role of the left hemisphere in motor control of touch: a functional magnetic resonance imaging analysis. *Brain Cogn.*, 49(1):96–101, 2002.

R. Borowsky, J. Loehr, C. Kelland Friesen, G. Kraushaar, A. Kingstone, and G.E. Sarty. Modularity and intersection of "what", "where" and "how" processing of visual stimuli: a new method of fMRI localization. *Brain Topogr*, 18(2):67–75, 2005.

E. Bullmore, M. Brammer, S.C. Williams, S. Rabe-Hesketh, N. Janot, A. David, J. Mellers, R. Howard, and P. Sham. Statistical methods of estimation and inference for functional MR image analysis. *Magn Reson Med*, 35(2):261–77, 1996.

A. Castles, T. Bates, and M. Coltheart. John Marshall and the developmental dyslexias. *Aphasiology,*, 20(9-11):871–892, 2006.

M. Coltheart. The genetics of learning to read. *J Res Read*, 29(1):124132, 2006.

M. Coltheart, K. Rastle, C. Perry, R. Langdon, and J. Ziegler. Drc: a dual route cascaded model of visual word recognition and reading aloud. *Psychol Rev*, 108 (1):204–56, 2001.

C. de Boor. *A practical guide to Splines*. Springer-Verlag, 1978.

E.A. DeYoe, P. Bandettini, J. Neitz, D. Miller, and P. Winans. Functional magnetic resonance imaging (fMRI) of the human brain. *J Neurosci Methods*, 54(2):171–87, 1994.

R.S.J Frackowiak, K.J. Friston, C.D. Frith, R.J. Dolan, and J.C. Mazziotta. *Human Brain Function*. Toronto: Academic Press, 1997.

K.J. Friston, P. Jezzard, and R. Turner. Analysis of functional MRI time-series. *Hum. Brain Mapp.*, 39:153–71, 1994.

K.J. Friston, A.P. Holmes, J.B. Poline, P.J. Grasby, S.C. Williams, R.S. Frackowiak, and R. Turner. Analysis of fMRI time-series revisited. *Neuroimage*, 2(1):173–81, 1995.

K.J. Friston, O Josephs, E. Zarahn, A.P. Holmes, S. Rouquette, and J. Poline. To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. *Neuroimage*, 12(2):196–208, 2000.

K.J. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–302, 2003.

R. Halari, T. Sharma, M. Hines, C. Andrew, A. Simmons, and V. Kumari. Comparable fMRI activity with differential behavioural performance on mental rotation and overt verbal fluency tasks in healthy men and women. *Exp Brain Res*, 169(1): 1–14, 2006.

D.A. Handwerker, J.M. Ollinger, and M. D'Esposito. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage*, 21(4):1639–51, 2004.

L. Lee, K. Friston, and B. Horwitz. Large-scale neural models and dynamic causal modelling. *Neuroimage*, 30(4):1243–54, 2006.

H.L. Liu, W.T. Liao, S.Y. Fang, T.C. Chu, and L.H. Tan. Correlation between temporal response of fMRI and fast reaction time in a language task. *Magn Reson Imaging*, 22(4):451–5, 2004.

T.T. Liu. Efficiency, power, and entropy in event-related fMRI with multiple trial types. Part ii: design of experiments. *Neuroimage*, 21(1):401–13, 2004.

T.T. Liu and L.R. Frank. Efficiency, power, and entropy in event-related fMRI with multiple trial types. part ii: design of experiments. *Neuroimage*, 21(1):387–400, 2004.

P. McDougall, R. Borowsky, G.E. MacKinnon, and S. Hymel. Process dissociation of sight vocabulary and phonetic decoding in reading: a new perspective on surface and phonological dyslexias. *Brain Lang*, 92(2):185–203, 2005.

R.S. Menon and S.G. Kim. Spatial and temporal limits in cognitive neuroimaging with fMRI. *Trends Cogn Sci*, 3(6):207–216, June 1999.

R.S. Menon, D.C. Luknowsky, and J.S. Gati. Mental chronometry using latency-resolved functional MRI. *Proc Natl Acad Sci U S A*, 95(18):10902–7, 1998.

W. Richter, P.M. Andersen, A.P. Georgopoulos, and S.G. Kim. Sequential activity in human motor areas during a delayed cued finger movement task studied by time-resolved fMRI. *Neuroreport*, 8(5):1257–61, 1997a.

W. Richter, K. Ugurbil, A. Georgopoulos, and S.G. Kim. Time-resolved fMRI of mental rotation. *Neuroreport*, 8(17):3697–702, 1997b.

W. Richter, R. Somorjai, R.and Summers, M. Jarmasz, R.S. Menon, J.S. Gati, A.P. Georgopoulos, C. Tegeler, K. Ugurbil, and S.G. Kim. Motor area activity during mental rotation studied by time-resolved single-trial fMRI. *J Cogn Neurosci*, 12 (3):310–20, 2000.

D. Ruppert, M.P. Wand, and R.J. Carrol. *Semiparametric regression*. Cambridge University Press, 2003.

G.E. Sarty and R. Borowsky. Functional MRI activation maps from empirically defined curve fitting. *Concepts in Magnetic Resonance Part B*, 24B(1):46–55, 2005.

V.A. Vakorin, R. Borowsky, and G.E. Sarty. Characterizing the fMRI response using Tikhonov regularization. *Statistics in Medicine, under review*, 2006.

E. Zarahn, G.K. Aguirre, and M. D'Esposito. Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *Neuroimage*, 5(3):179–97, 1997.

Chapter 6

Conclusion

The purpose of this work was to extend the applicability of functional magnetic resonance imaging for gaining futher insight into human brain function by relating behavioral measures to underlying brain activation. Specifically, we aimed (i) to explore the mathematical relationships between fMRI blood oxygen level dependent (BOLD) signals and evoked neural activity, and (ii) to extract more behavioral information from the dynamics of the hemodynamic responses.

Although the thesis is composed of self-contained chapters, the parts are intrinsically interconnected with each other. In Chapter 2 we considered the hemodynamic inverse problem (*i.e.* the problem of extracting neural activity underlying the hemodynamic response). Our aim was to propose techniques that can reproduce the global solution of the hemodynamic inverse problem, avoiding spurious local minima. We found that the dynamics of neural activity can be robustly reconstructed from observed hemodynamic responses when the physiological link is specified by the so called expanded "balloon" model. In addition, we showed that the theoretical solution to the hemodynamic inverse problem based on the balloon model is on/off dynamics of neural activity.

In Chapter 3 we explored the possibility of recovering hidden information from

fMRI data through solving the hemodynamic inverse problem, switching the focus of our study to the neural level. We applied the techniques proposed in Chapter 2, coupled with signal averaging and smoothing at the neural level. This approach was illustrated by example with a functional network model of an object interaction decision task, based on a dorsal-ventral model of visual information processing. At the hemodynamic level, we could not reveal any latency in activation. Going from the hemodynamic level down to the neural level, we were able to find statistically significant delays in information processing between some of the functional units of the model. In addition, the estimates of connectivity and chronometry, obtained at the neural level, provide support for the hypothesis that it is the dorsal pathway which is dominant in performing a task of word reading and deciding how to interact with its referent object.

Chapter 4 probes the issue of estimating the true shape of the discrete and noisy hemodynamic response. Specifically, we considered the problem of the optimal smoothing of averaged fMRI responses for repeated block design experiments. Simultaneous averaging and smoothing was performed using functional data analysis with Tikhonov regularization. We found that the autocorrelation structure of residuals can be improved with the appropriate selection of the regularization parameter. Also, generalized cross-validation does not guarantee the best correlation structure of the residuals as determined by a white noise test. Finally, we concluded that caution should be taken with blind smoothing of fMRI data as there is no guarantee that the fMRI time series would not be oversmoothed.

The ultimate goal of the smoothing techniques considered in Chapter 4 is to

facilitate computations of temporal characteristics of the BOLD response in order to verify previously reported correlations between the hemodynamics and behavioral measures. In Chapter 5 we applied the methods developed in Chapter 4 to compute maps of several parameters of the BOLD responses in Broca's area during a word identification task. The results revealed that behavioral reaction time was correlated with functional BOLD width for naming pseudohomophones, but not for regular words, irregular words and non-words. Further, the results provided additional support for an independent dual-route model of reading, extending the independent nature of lexical and sub-lexical systems from the behavioral level, reported previously in terms of reading accuracy, to the hemodynamic level in terms of BOLD width, signal intensity and initial slope.

A strong assumption of the approach considered in Chapter 2 is the validity of the biophysical model of the transduction of neural activity into the BOLD signal. However, the model is far from being complete, especially taking into account the unclear nature of the neurovascular coupling. A more detailed analysis of the physiological link between neural activity and BOLD response in the future would be required for more accurate predictions. Several modifications of the balloon model, which can simplify or complicate the original model, have been recently proposed in the literature. A comparative study of these models, based on multimodal measurements at the neural and hemodynamic level, is a direct way to test and refine mathematical relationships between hemodynamic responses and underlying neural activity. This approach may incorporate mathematical analysis with the goal to obtain more accurate estimates of the biophysical parameters, which control the

transitional dynamics of the physiological variables in question. Such studies can potentially include parameter sensitivity analysis, which is crucial for estimating the delay of BOLD response relative to the onset of neural activity, and consequently for tracing relative delays in activation between different functional units of a network model.

Chapter 3 presents an attempt to decompose an object interaction decision task into a sequence of activation associated with visual word/object recognition, interaction semantics and speech production, by recovering information from fMRI data via solving the hemodynamic inverse problem. Although we were able to find statistically significant latencies, which, from the qualitative point of view, are in concordance with the previously reported results from event-related potential (ERP) language-related studies, latency in activation was not under direct experimental control. An avenue for future work would involve an experiment designed to activate distinct regions (for example, left and right hemifields) with variable delay in stimulus presentation. A next step would be to determine the degree to which the techniques, developed to solve the hemodynamic inverse problem and to make inferences at the neural level, are capable of recovering the onset latency that is controlled experimentally.

In addition, two points are worth mentioning, regarding the methodology we followed in Chapter 3 to produce the BOLD responses associated with functional modules via averaging fMRI times series of the activated voxels within modules. One concern reflects physiologically determined variability of BOLD responses, which is not related to underlying neural activity. As delay between BOLD responses and

their neural substrates can vary within and across regions by up to ± 2 seconds, averaging within functional modules can potentially eliminate relevant temporal information. Another concern is that activated voxels attributed to one functional module can belong to separate slices, characterized by different acquisition time. A possible correction is to consider a more detailed network model, with subunits of the previously masked functional units, the voxels of which are characterized by smaller dispersion values of the temporal parameters of the hemodynamic responses.

Chapter 4 considers possible criteria for selecting a regularization parameter which controls the amount of smoothing imposed on fMRI time series required for further analysis. The problem is to decide which criterion is best suited for choosing the "right" amount of smoothing. In an attempt to compare different smoothing criteria, in Chapter 4 we proposed to use our expectations about the shape of the hemodynamic response, but such an approach can be compromised by the variability of the hemodynamic responses, reported to be varying not only across subjects but also from voxel to voxel and from task to task. Another possible way to address this issue is by a relative comparison of hemodynamic responses rather than in estimating the shape of the BOLD response in an absolute way. Specifically, a choice in favor of one regularization parameter criterion or another can be motivated by the ability of this criterion to better reproduce timing differences between hemodynamic signals measured in an experiment with controllable stimulus latencies.

As illustrated in Chapter 5, the problem of estimating temporal and spacial parameters of the hemodynamic response requires some kind of smoothing. As a result, the estimated values of the BOLD response parameters depend on some fac-

tors characterizing the amount of smoothing applied. Specifically, such factors like the number of basis functions, the knot sequence, the degree of polynomials and the type of basis functions can affect the estimates of BOLD parameters, which might be crucial for further correlation analysis. Parameter sensitivity analyses would be desirable to ensure that methodological aspects do not effect final results regarding the relationship between behavioral and functional measures.

Finally, we would like to emphasize that the proposed techniques, in general, are computationally expensive. However, this shortcoming is compensated by the capacity to extract additional information from hemodynamic responses, extending applicability of fMRI-based methods for studying brain function.