fMRI activation maps based on the NN-ARx model

[1]Riera J., [2]Bosch J., [3]Yamashita O., [1]Kawashima R., [4]Sadato N., [5]Okada T., [3]Ozaki T.

[1]NICHe, Tohoku University, Sendai

[2]Cuban Neuroscience Center, Havana

[3]The Institute of Statistical Mathematics, Tokyo

[4]National Institute for Physiological Science, Okazaki

[5]Institute of Biomedical Research and Innovation, Kobe

Correspondence to: Dr. Jorge J. Riera

Advanced Science and Technology of Materials
NICHe, Tohoku University
Aoba 10, Aramaki, Aobaku, Sendai 980-8579, JAPAN
TEL/FAX +81-22-217-4088
Office email: riera@idac.tohoku.ac.jp
Personal email: riera_jorge@hotmail.com

Abstract

The most significant progresses in the understanding of human brain functions have been possible due to the use of fMRI, which, when used in combination with other standard neuroimaging techniques (i.e. EEG), provides researchers with a potential tool to elucidate many biophysical principles, established previously by means of animal comparative studies. However, to date most of the methods proposed in the literature seeking fMRI signs have been limited to the use of a top-down data analysis approach, thus ignoring a pool of physiological facts. In spite of the important contributions achieved by applying these methods to actual data, there is a disproportionate gap between theoretical models and data-analysis strategies while trying to focus on several new prospects, like for example fMRI/EEG data fusion, causality/connectivity patterns and nonlinear BOLD signal dynamics. In this paper, we propose a new approach which will allow many of the above mentioned hot topics to be addressed in the near future with an underlying interpretability based on bottom-up modeling. In particular, the θ-*MAP* presented in the paper to test brain activation corresponds very well with the standardized T-test of the SPM99 toolbox. Additionally, a new Impulse Response Function (IRF) has been formulated, directly related to the well-established concept of the *hemodynamics response function*. The model uses not only the information contained in the signal but also that in the structure of the background noise to simultaneously estimate the IRF and the autocorrelation function by using an autoregressive model with a filtered Poisson process driving the dynamics. The short-range contributions of voxels within the near-neighborhood are also included, and the potential drift was characterized by a polynomial series. Since our model originated from an immediate extension of the hemodynamics approach (Friston et al. 2000a), a natural interpretability of the results is feasible.

Introduction

Functional Magnetic Resonance Imaging (fMRI) is a very useful technique to study brain functions while the subjects are involved in the performance of sensory, motor, or cognitive tasks. In many experimental conditions, the observed Blood Oxygenation Level Dependent (BOLD) signals reveal highly nonlinear dynamics (Berns et al. 1999, Friston et al. 1998b, Huettel and McCarthy 2000 and 2001, Birn et al. 2001). Recently, an ordinary differential equation system (i.e. the Balloon model) has been proposed to explain the hemodynamic changes on the basis of the mechanically compelling model of an expandable venous compartment (Buxton et al. 1998) and the standard Windkessel theory (Mandeville et al. 1999). The original model has been extended by Friston et al. (2000a) to include a linear interaction between synaptic activity (or electrophysiology) and the micro-vascular control system. The model, known as the *"hemodynamics approach"*, is in accordance with many recent physiological findings (Magistretti and Pellerin 1999; Iadecola 2002). However, to date, few reports have been published about the application of this approach to actual fMRI data. The most significant studies have focused on the use of Volterra Kernel expansion (Friston 2002) and Local Linearization (LL) filter (Riera et al. 2004). Despite the high computational cost involved when using optimization strategies to estimate the parameters of nonlinear approaches, the imminent interpretability of these parameters makes it worthwhile. Unfortunately, only in particular cases could the non-linearities of BOLD signals be explored in selected brain regions based on bottom-up modeling; hence, as yet we are not even close to the point where they can be used in either clinical or research studies focusing on testing brain activity.

Even thought the nonlinear numerical schemes are still at the development stage, there are many "*linear models*", constituting a first order approximation of the BOLD

signal dynamics, which have emerged in the last ten years. However, it is our belief that linear models must be introduced in a way that makes generalizations feasible in order to consider more complex dynamics. The authors consider it necessary to make a brief review of the most relevant aspects of linear models proposed to date in the literature in order to lead the readers to a natural understanding of the new approach. The pioneer paper of Friston et al. (1994) represents the foundations of this work, having set up the basic idea of using a linear convolution of the now popularized Hemodynamics Response Function (HRF) with the neuronal process, which corresponds to synaptic activation. There are two physiological phenomena underlying the BOLD response at the level of the synapses that contribute to the neuronal process proposed by Friston et al. (1994): the evoked transient and the uncorrelated intrinsic activity. This model instinctively originated from the fact that observed Auto-Correlations Function (ACF) have two components: one of them the effect of being phase-locked to the stimulus onset of the neuronal evoked transients and the other due to spontaneous neuronal activity. In an early work, Weisskoff et al. (1993) reported similar characteristics of the power spectral density of the BOLD signals. At that time, there was some debate as to whether the correlation in the BOLD signal has a physiological or instrumental source, or even if it is just the result of mixing both types of sources.

The implicit merit of the suggestion made by Friston et al. (1994) is the common vascular etiology underlying the genesis of the predictable hemodynamics (i.e. deterministic response induced by the stimulus) and the characteristics of its fluctuations (i.e ACF). The successively proposed linear models have disregarded that physiological conjecture, maybe in an attempt to provide a high degree of freedom to the mathematical constructs, thus guaranteeing a more accurate portrayal of the

temporal dynamics of fMRI. These unconstrained linear models assume that an unspecified correlated noise additively polluted a deterministic response produced by the convolution of the HRF with a predefined binary stimulus sequence. The use of fixed or jittered Inter Stimulus Interval (ISI) in the stimulus sequence has been proposed in the literature (see Dale 1999 for comparisons using the **estimator efficiency** as a new selection criterion of paradigm adequacy). In general, this sequence can be considered as a Poisson random process defined by the experimental design. In the last ten years, different methods have been proposed to accommodate fMRI data into that extended version of the original Friston et al. (1994) linear model. In this sense, the most recent works have been concerned with the simultaneous estimation of the HRF and the ACF of the additive noise (Bullmore et al. 1996, Locascio et al. 1997, Dale 1999, Kruggel and von Cramon 1999, Burock and Dale 2000, Woolrich et al. 2001, Purdon et al. 2001, Worsley et al. 2002, Katanoda et al. 2002, Friston and Penny 2003). Paradoxically, the main motivation for estimating the ACF has been the fact that by its a priori knowledge: a) the Maximum Likelihood (ML) estimator of the supposedly deterministic HRF can be corrected by the temporal variance-covariance scales of noise (Kruggel and von Cramon 1999, Katanoda et al. 2002); and b) a considerable reduction of the bias in any statistical test used to hypothesize about the effect of the contrast will be achieved (Worsley and Friston 1995, Worsley et al. 1997, Friston et al. 2000b, Marchini and Smith 2003). The consideration of both facts in the fMRI analysis using a unified formalism has been also presented (Burock and Dale 2000, Woolrich et al. 2001, Friston and Penny 2003). In a slightly different approach, the inclusion of additional information about the temporal characteristic of HRF in both parametric and non-parametric schemes has enhanced linear models. These methods range from those of pioneers who constrained

the HRF, forcing it to belong to a certain space expandable by basic functions (Lange and Zeger 1997, Rajapakse et al. 1998, Friston et al. 1998a, Josephs et al. 1997) to the most recent methods that introduce smoothness criterion in the a priori probability function of the HRF (Goutte et al. 2000, Carew et al. 2003, Marrelec et al. 2003). The problem with this approach is that, in both the parametric and non-parametric cases, the use of constraints (or forced assumptions) on the HRF could lead to a severe mismatch between the anticipated deterministic response and the actual background noise, which, from the conjecture proposed by Friston et al. (1994), may share a similar vascular mechanism of genesis (i.e. see Riera et al. 2004 for a discussion about the impact on the BOLD signal produced by deterministic/noisy inputs in the hemodynamics approach).

However, the common and critical intuitive idea underlying those methods previously proposed in the literature to model the hemodynamic changes linearly does not correspond with the essential physiological and physical principles involved in the genesis of BOLD signal. Rather, it constitutes an ad-hoc tactic introduced to bypass a more complex general identification problem. It is well known that two connected stages of functioning with different temporal scales coexist at a microscopic level for the BOLD dynamics induced by a stimulus sequence: the quasi-linear spatial-temporal integration accomplished at the neuron-astrocyte unit and a low-pass nonlinear filter acting in the hemodynamic/metabolic order at the micro-vascular building block. Therefore, in our opinion, it is important that from the very beginning, the focus of every aspect of the methodology is on the formalism of the time series analysis since it permits us to generalize at a later time any higher order of complexity in an appropriate framework. In this paper, a first order linear approach is presented which separates the contributions to BOLD signals coming from both the *cause* – the

electro-chemical phenomena at the dendrite trees, and the *effect* – the vascular response. The model comprises both a linear filter and an AutoRegressive AR(p) component, the former simulating several stages of the spatial-temporal integration process (i.e. neurotransmitters migration from synaptic cleft, transport phenomena at the neuron-astrocytes juncture, and electrotonic propagation of post-synaptic potentials, etc) that relates the evoked transient to the Poisson stimulus sequence. The latter accounts for the hemodynamic variations being driven by the neuronal process as an additive random force. This AR model, hereafter denoted as ARx(p,q) due to the use of an exogenous variable, will not only permit an estimation of the hemodynamical response by the use of a new concept of *"Impulse Response Function"* (IRF) but will also make it possible to have a natural description of the ACF implicitly found in BOLD signals. In this sense, the attributes in the micro-vascular subsystem will affect both the final deterministic response and the out-coming noise. Because of a causal relationship, the authors consider the information contained in the structure of noise of the BOLD signal to estimate the IRF extremely valuable. The concept of "*innovation*" in the general "Box-Jenkins representation" is used, which could dramatically change the existing course of fMRI analysis.

This model, surprisingly, allowed us to detect those regions directly related to the stimulus with a very low spatial localization error. The IRF and the ACF were very well characterized in each voxel using little data, which enabled the time the experiment takes to be reduced considerably (lengthy experimental time is a handicap in fMRI study). Additionally, the model was extended to include contributions from the near-neighborhood of each voxel by introducing the influence of past short-range interactions as suggested earlier by Purdon et al. (2001) and Katanoda et al. (2002). The performance of the model, henceforth named Near-Neighborhood

AutoRegressive with exogenous variable (NN-ARx), was evaluated from synthetic data created by applying the LL scheme (Riera et al. 2004) to discretizing the original hemodynamics approach. The Kolmogorov-Smirnov test was used to evaluate how significantly the histogram of the innovation process departed from the gaussian distribution, a well fitting criterion of the model. A corrected version of the Akaike Information Criterion (AIC) (Hurvich and Tsay 1989), the so-called AICc, was introduced for model selection since it endows us with a rule to estimate the order of NN-ARx (length in the influence of the past for the AR and the Poisson filter), the order for the nonlinear drift component, the delay of the signal respect to the stimulus and the retard in the near-neighborhood component due to physical distances. The AICc implicitly gauges the complexity of the data/model, avoiding overestimation. An inverse Laplacian pre-filtering method was applied in advance to the raw BOLD data to whiten any spatial correlation unfavorably introduced by the application of gaussian kernels in the fMRI recording systems. The whole method was applied to two experimental situations using block design paradigms (i.e. motor and visual stimuli). In order to assess the robustness of the methodology, different MRI recording systems and parameters were used in both experiments. The most significant areas obtained from the application of the proposed method (a simple cluster classification made by thresholding the hot-spots) were compared with those reported after analyzing the data with statistical parametric mapping software (SPM99 toolbox, Welcome Department of Cognitive Neurology, London, UK).

Methods

Experiments design

In order to explore the robustness of the method to experimental manipulations, the two sets of BOLD data used in the paper were obtained under very different physiological and recording conditions.

Visual paradigm: A 3-T scanner (VP, General Electric, Milwaukee, WI) was used in this study. Ten normal volunteers (5 males and 5 females) aged 25-43 years were used in the visual paradigm consisting of 3 blocks of 30 Secs checkerboard visual stimulus and 30 Secs of control condition (starting from task condition) (see Fig. 5 bottom). During the task condition, the checkerboard was intermittently presented at a frequency of 8 Hz. Tight but comfortable foam padding was placed around the subject's head to minimize head movement.

fMRI parameters: Inter-scan interval $TR = 3$ Secs. Each volume consisted of 36 slices from the bottom to the top of the head, with a voxel size of 3.44 x 3.44 mm in plane, a slice thickness of 3.5 mm and a 0.5 mm gap covering the whole brain. T2-weighted, gradient echo, echo planar imaging (EPI) sequences. ($TE = 30$ mSecs, $FOV = 22$ cm)

Parameters of scanner for anatomical reference: T2-weighted, 2D-fast spin echo sequence (with parameters of $FA = 90$ degree, $TR = 6000$ mSecs and $TE = 70$ mSecs) consisting of 112 trans-axial slices, with slice thickness 1.5 mm, and pixel size was 0.859 x 0.859 mm.

Motor paradigm: A 1.5-T scanner (Vision, Siemens, Erlangen, Germany) was used in this study. Five right-handed, normal volunteers (3 males and 2 females) aged 24-

37 years were used in the motor paradigm consisting of 9 blocks of 60 Secs moving conditions and a 60 Secs resting condition (starting from resting condition) (see Fig. 7 bottom). The subjects were asked by visual cues (at regularly spaced intervals with a frequency of 1.6 Hz) to perform right hand movement tasks. During the moving condition, a small circle at the center of the screen was used as a cue (lasting for 200 mSecs) indicating the subject should close its hand and a cross indicating to open it. Each subject's head was fixed using ear fixation blocks.

fMRI parameters: Inter-scan interval $TR = 1.2$ Secs. Each volume consisted of 8 slices from top to bottom of the head, with a voxel size of 3 x 3 mm in plane, a slice thickness of 10 mm and with a 5 mm gap covering the whole brain. T2-weighted, gradient-echo, echo-planer imaging (EPI) sequences ($TE = 60$ mSecs, $FA = 90$ degrees).

Parameters of scanner for anatomical reference: Spoiled gradient-echo sequence (recovery time $TR = 9.7$ mSecs, echo time $TE = 4$ mSecs, $FA = 12$ degrees) consisting of 96 slices with a voxel size of 1.25 x 0.9 x 1.92 mm.

Data pre-processing

In both paradigms, the individual fMRI images were realigned to remove movement-related artifacts, and the slice timing was adjusted to that of the middle slice. The anatomical and fMRI images were co-registered and spatially normalized to the Talairach coordinate system using both linear and nonlinear parameters. The raw fMRI data was spatially whitened using the inverse of the laplacian operator to eliminate any nuisance spatial autocorrelation introduced by the previous usage of

volumetric gaussian kernels (see Appendix I). Henceforth, the symbol y_t^v will be used to identify the scan "*t*" of preprocessed BOLD data at the *v*-th voxel.

Theoretical Model

The consequences of the convolution model

The original model by Friston et al. (1994) established that a BOLD signal $y^v(t)$ at a particular voxel "*v*" and time "*t*" could be represented by a convolution of the neuronal process $\xi^v(t)$ (i.e. at the level of synapses) with the effective voxel-related HRF $h^v(\tau)$.

$$y^v(t) = \int_0^\infty h^v(\tau)\xi^v(t-\tau)\,d\tau \qquad (1)$$

It was proposed that the neuronal process comprised a deterministic evoked transient $e^v(t)$ and uncorrelated fast intrinsic activity $\varepsilon^v(t)$. The "*Wold decomposition*" of an AR Moving-Average (ARMA) model $y_t^v = \mathrm{H}^v(B)\xi_t^v$, with the infinite lag polynomial defined by $\mathrm{H}^v(B) = \sum_{k=0}^{\infty} h_k^v B^k$ and weights h_k^v, can be interpreted as a discretized version of the convolution operator. The symbol B denotes the backshift (or lag) operator. The weights must satisfy the property $\sum_{k=0}^{\infty}\left(h_k^v\right)^2 < \infty \;\; \forall_v$ (see Appendix II for details). This model assumes that neural process distributes $\xi^v \sim N\left(e^v, \sigma_v^2\right)$, with voxel-dependent standard deviation σ_v.

In general, providing certain conditions are met, the Wold decomposition can be approximated by the fraction of finite polynomials $\mathrm{H}^v(B) = \dfrac{\Psi^v(B)}{\Phi^v(B)}$. In this paper, the coefficients ψ_k^v of polynomial $\Psi^v(B)$ will be set to zero, in order to collapse the

whole hemodynamics in the p-order characteristic polynomial $\Phi^v(B) = 1 - \sum_{k=1}^{p} \phi_k^v B^k$.

An $\mathrm{AR}(p)$ will be invertible if all the zeros of $\Phi^v(B)$ lie outside the unit circle (Brockwell and Davis, 1987). This condition will ensure that the original convolution model (1) is equivalent to a "*stationary*" (comparable to a causality condition) $\mathrm{AR}(p)$, with the neuronal process ξ_t^v representing an additive random force driving the system, far from being a white noise.

$$y_t^v = \sum_{k=1}^{p} \phi_k^v y_{t-k}^v + \overbrace{e_t^v + \varepsilon_t^v}^{\xi_t^v} \qquad (2)$$

Therefore, the $\mathrm{AR}(p)$ (2) is externally perturbed by an unknown deterministic input (i.e. the evoked transient $e^v(t)$), which we are also interested in estimating. A recursive relationship can be used to compute the coefficients h_k^v $(k = 1, \cdots, \infty)$ from given values of ϕ_k^v (see equation II-6 in Appendix II).

The classical linear model

In recent years, the original formula (1) has been lightly modified; the voxel-dependent evoked transient has been directly associated with a *common* stimulus sequence $s(t)$ that *finitely*-convolves with the HRF of each voxel. Additionally, an unknown noise component ζ_t^v has been included in an attempt to capture the most significant characteristics of the observed ACF (see equation 3 below). In this approach, the hemodynamic deterministic response (i.e. BOLD signal) and the ACF do not share a similar etiology as in the original Friston et al. (1994) model, and the fact that $\varepsilon^v(t)$ may originate at the level of synapses and, if so, will be thus colored by the vascular filter has been disregarded. In several recent papers, parametric and

non-parametric methods have been used to estimate the ACF of the noise component ζ_t^v at the same time that the HRF is tailored to the BOLD data.

$$y_t^v = \sum_{k=0}^{T} h_k^v s_{t-k} + \zeta_t^v \qquad (3)$$

In a vector representation, the linear model (3) can be written as:

$$\mathbf{y}^v = A\boldsymbol{\gamma}^v + \boldsymbol{\zeta}^v \qquad (4)$$

The vector $\mathbf{y}^v = \left(y_1^v, \cdots, y_N^v\right)^t$ summarizes the BOLD signals time series (i.e. N scans) observed in the *v*-th voxel and vector $\boldsymbol{\gamma}^v = \left(h_0^v, \cdots, h_T^v\right)^t$ comprises the truncated HRF. The $(N \times T)$ experimental design matrix A is constructed from the pre-defined ISI. Each stimulus is approximated by an instantaneous delta Dirac (i.e. a Poisson stochastic process). The vector noise component is assumed to distribute $\boldsymbol{\zeta}^v \sim N\left(0, \Sigma^v\right)$, with unknown voxel-dependent variance-covariance matrix Σ^v.

An ill-posed inverse problem originates while trying to estimate the HRF and Σ^v simultaneously from data $\mathbf{y}^v$, especially due to the very sparse and particular structure of the design matrix A. Dale (1999) has reported an improvement in the HRF estimation using a randomly distributed ISI, which guarantees a considerable increase of the rank of the matrix $A^t A$. The use of the Bayesian framework has allowed the introduction of a priori information about the HRF (i.e. smoothness criterion), which yields to Maximum A Posteriori (MAP) estimators (Goutte et al. 2000, Carew et al. 2003, Marrelec et al. 2003) (i.e. the use of a priori probability function $P\left(\boldsymbol{\gamma}^v / Q^v\right) \sim \exp\left(-\left(\boldsymbol{\gamma}^v\right)^t \left(Q^v\right)^{-1} \boldsymbol{\gamma}^v / 2\right) \Big/ \sqrt{\left|Q^v\right|}$). In this line of work, the HRF is forced to be temporally smooth using a discrete model of the second order derivative in the inverse of the a priori variance-covariance matrix Q^v, which warrants a stable

reconstruction of solutions in case of very ill-posed identification problems. Parametric models of Q^v are also possible, where the voxel-dependent hyperparameters can be estimated from data using Bayesian arguments in a second level hierarchical model. In the general case, the MAP estimator of vector $\hat{\boldsymbol{\gamma}}^v$ is given by:

$$\hat{\boldsymbol{\gamma}}^v = \left(A^t \left(\Sigma^v \right)^{-1} A + \left(Q^v \right)^{-1} \right)^{-1} A^t \left(\Sigma^v \right)^{-1} \mathbf{y}^v \qquad (5)$$

Also, the BOLD signal time series is assumed to be a stationary stochastic process; therefore, the estimation of the full variance-covariance matrix Σ^v is unsuitable since it has a Toeplitz structure. To overcome this, AR models have been used in the last few years, with pre-defined hyperparameters to be estimated. However, in our opinion, model (1) must be used in its original form; and particular approaches for the evoked transient component should be proposed on the basis of physiological interpretability.

The NN-ARx model and the hemodynamics approach

The original Balloon model (Buxton et al. 1998, Mandeville et al. 1999) was extended by Friston et al. (2000a) to include a linear interaction between the synaptic activity and the micro-vascular control system (i.e. the hemodynamics approach). Fig 1 (**a**) shows a simple scheme of such a dynamic system after being generalized to include a physiological noise (Riera et al 2004) added to a linearly filtered input and instrumental errors. The nonlinear block (dot-line box) consists of a dynamics subsystem (i.e. hemodynamics approach) and a static/nonlinear function $g(\cdot)$, which represents the observation equation and explicitly depends on some of the state-variables of the hemodynamics approach (Buxton et al. 1998). The linear filter, which emulates the spatial-temporal integration at the level of the neuron-astrocyte unit, transforms a Poisson process representing the stimulus sequence into the evoked

transient e_t^v that combines with a physiological noise ε_t^v to create the final neuronal process ξ_t^v. The neuronal process initiates the nonlinear hemodynamics subsystem through a flow-inducing variable (at the present, it is recognized the role played by nitric oxide and sphincters in that vascular control feed forward mechanism), the outputs of which are the state-variables. At the last stage, the BOLD signal y_t^v originates from applying the static/nonlinear function $g(\cdot)$ to the state-variables, finally contaminated by instrumental errors η_t^v. Fig. 1 (**b**) shows a symbolic linearization of the nonlinear block (i.e. dot-line box). It is assumed that the linear filter relating the neuronal process to the common stimulus sequence s_t has a finite parametric form (see equation 6). The coefficients θ_k^v can be interpreted as the strength or magnitude of the synaptic activity in the *v*-th voxel induced by the *(k+d)*-th temporal lag of s_t (i.e. the Poisson process). The delay *d* is introduced to consider either quasi-instantaneous or retarded synaptic responses, and it must be estimated from data. In general, in the original hemodynamics approach the instrumental errors η_t^v were assumed to be a pure white noise process. This, in addition to the linearization of the nonlinear block, justifies the collapse of the variance of η_t^v into the statistical moments of the physiological noise ε_t^v.

$$e_t^v = \sum_{k=0}^{r} \theta_k^v s_{t-k-d} \tag{6}$$

Hence, the equation (2) can be interpreted as an AR model, with exogenous variables s_t (see Appendix II). The contribution of near-neighborhood dynamics $\boldsymbol{\xi}_{t-\Delta}^v = \left\{ y_{t-\Delta}^{v'}, v' \in \Omega_v \right\}$ is also included, where the vector $\boldsymbol{X}^v = \left\{ \chi_v^{v'}, v' \in \Omega_v \right\}$ summarizes the anisotropic properties of the local vascular correlations (Purdon et al.

2001, Katanoda et al. 2002). The magnitude Δ represents a delay that may occur due to the effect of the mean distance between the voxel of interest and those within its neighborhood Ω_v.

The nuisance effects are included in the model by using a voxel-dependent nonlinear potential drift $\mu_t^v = \sum_{k=0}^{\delta} \gamma_k^v t^k$, which is represented by a polynomial series. The parameters of the model $\Xi^v = \left\{ \phi_k^v, \theta_k^v, \gamma_k^v, \sigma_v, \boldsymbol{X}^v \right\}$ must be estimated for each voxel from the BOLD signals (see Appendix III for details). The model selection consists of determining both the model orders and the delays. These magnitudes, related to the complexity of the dynamics, are denominated global parameters and are comprised in the vector $\Lambda = (p, r, d, \Delta, \delta)$.

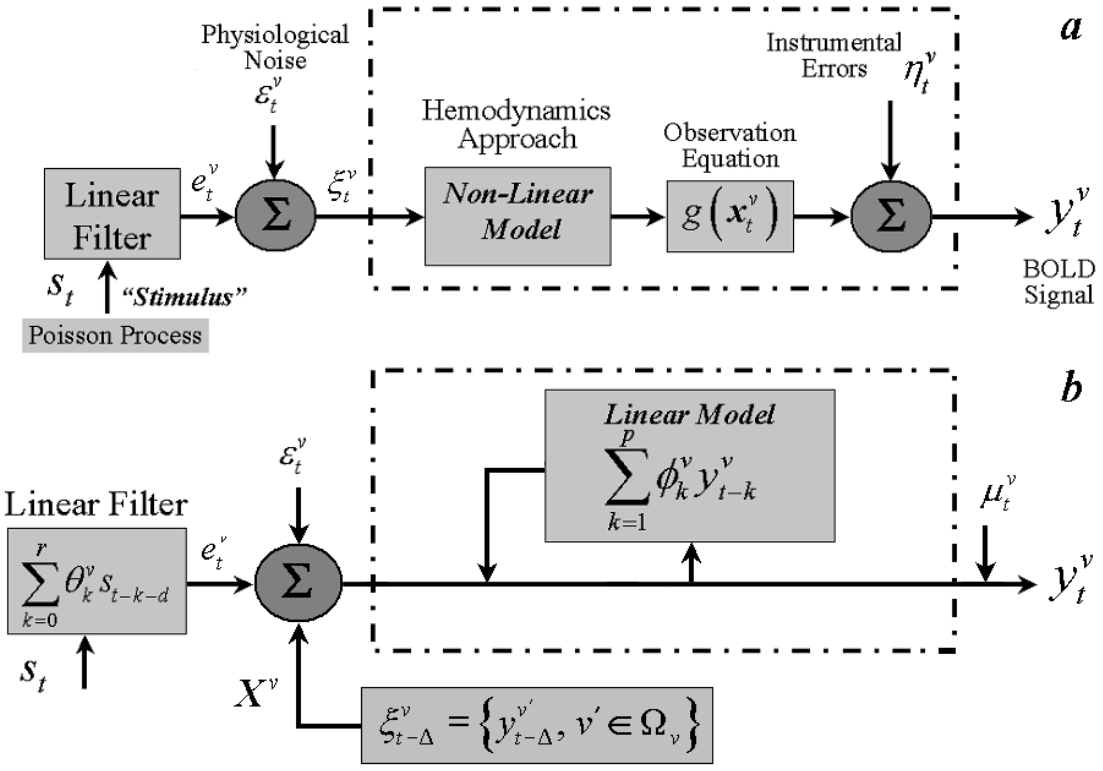

Fig. 1. *Schematic diagrams.* ***(a)*** *A system consisting of a linear filter and a nonlinear block (dot-line box), which is comprised of the hemodynamics approach and a*

*static/nonlinear observation equation. The whole system is extended to include both instrumental error and physiological noise. (**b**) The NN-ARx linear model, which is comprised of a linearization of the nonlinear block (i.e. dot-line box), a parametric form for the linear filter, contributions from the near-neighborhood* Ω_v *dynamics (with anisotropic factors* $\boldsymbol{X}^v$*), the physiological noise and the potential drift term.*

The NN-ARx model (7) includes four terms: the potential drift, the hemodynamics linear model, the local dynamics contributions, the weighted stimulus sequence by synaptic effectiveness and the additional diffusion term (i.e. the white noise ε_t^v).

$$y_t^v = \mu_t^v + \sum_{k=1}^{p} \phi_k^v y_{t-k}^v + \boldsymbol{X}^v \boldsymbol{\xi}_{t-\Delta}^v + \sum_{k=0}^{r} \theta_k^v s_{t-k-d} + \varepsilon_t^v \tag{7}$$

The HRF (associate with $\mathrm{H}^v(B)$), as originally defined in the literature, and the genuine IRF (associate with polynomial $\Gamma^v(B) = \sum_{k=0}^{\infty} g_k^v B^k$) for the stimulus sequence will explicitly depend on the model parameters ϕ_k^v and θ_k^v. It is proper to clarify that to obtain such relationships, the contribution of local dynamics and potential drift terms are handled as instantaneous and deterministic inputs to the system; hence, they will be ignored. The coefficients of the HRF and the IRF can be obtained by recursive relationships (see Appendix II for the general theory of ARMA models with/without exogenous variables). In the particular case of model (7), these standard relationships must be slightly modified to include the delay d, in such a case $\overline{\Psi}(B) = [1 \quad \Theta(B)B^d]$; hence, the coefficients of the HRF and the IRF will satisfy the following relationships:

$$h_k - \sum_{k'=1}^{\min(k,p)} \phi_{k'} h_{k-k'} = 0 \tag{8}$$

with $h_0 = 1$

$$\begin{aligned} g_k - \sum_{k'=1}^{\min(k,p)} \phi_{k'} g_{k-k'} &= 0 \qquad k < d \\ g_k - \sum_{k'=1}^{\min(k,p)} \phi_{k'} g_{k-k'} &= \theta_{k-d} \qquad k \geq d \end{aligned} \tag{9}$$

with $g_0 = \begin{cases} \theta_0 & d = 0 \\ 0 & d \neq 0 \end{cases}$

Equation (8) is equivalent to (II-6) (because $\psi_k^v = 0$ in our model). Furthermore, for the particular case of insignificant delay in the stimulus (i.e. $d = 0$), the equations (9) and (II-12) are analogous. Additionally, the weights h_k^v will be involved in computing the ACF of the physiological noise. The new auto-correlated noise $\mathrm{H}^v(B)\varepsilon_t^v$ will have an ACF $R^v(\tau)$ defined by equation (II-7) in Appendix II.

In the model, the magnitude $\boldsymbol{\theta}^v = \sum_{k=0}^{q} \theta_k^v$, henceforth referred to as "θ-*MAP*", determines the spatial distribution of the brain synaptic sensitivity to the stimulus sequence.

The subroutines used in this paper are available in MATLAB 5.3 code for any reproducible research: 1) the LL method to discretize hemodynamics approach, 2) the ML optimization algorithm for estimating the parameters of the model, and 3) the AICc introduced in order to carry out model selectivity.

Results

Synthetic data

In this section, BOLD synthetic data was created using the original nonlinear hemodynamics approach (Fig. 1 (**a**)) to evaluate the consistency of the method. The hemodynamic approach can be mathematically formulated by a nonlinear and non-autonomous Stochastic Differential Equations (SDE) system that relates a states

vector $\boldsymbol{x}(t)$ to the neuronal synaptic activity $u(t)$ (see Fig 1 and equation (1) in Riera et al. 2004). The LL method (Jimenez and Ozaki 2003) uses random measures to integrate the SDE in the vicinity of discretely and regularly distributed time instants assuming a local piecewise linearity (see equation 6 in Riera et al. 2004). Therefore, the LL formalism permits the conversion of a SDE system into a vector states equation with a background gaussian noise, where a stable reconstruction of the trajectories of the states vector is obtained by a one-step straightforward prediction (i.e. a nonlinear AR model). Finally, the BOLD signal relates to the states vector $\boldsymbol{x}_t$ by a nonlinear observation equation (Buxton et al. 1998).

The method proposed in this paper was applied to the simulated data to fit model parameters $\Xi^v = \{\phi_k^v, \theta_k^v, \gamma_k^v, \sigma_v\}$. Note that in this particular case, the contribution of local dynamics was not included in the model (7). The IRF and ACF were estimated simultaneously using the recursive relationships (8) and (9) in combination with the explicit equation (II-7). In order to estimate model complexity, the AICc was minimized with respect to global parameters $\Lambda = \{p, q, d, \delta\}$.

The hemodynamics approach generates two distinctive signaling modes: damped oscillations or an exponential decay behavior. The mean transit time in the post-capillary venous compartment has been interpreted for steady-state conditions as the time constant of an equivalent analogical RC circuit (i.e. Windkessel theory). Therefore, changes in this parameter will produce alterations in the *modus operandi* of the micro-vascular control system, slowing down its dynamics, and therefore avoiding oscillations in the IRF and ACF when the parameter is increased. Fig. 2 shows the results of applying our method to BOLD signals generated with a transit time of 0.1 Secs. In Fig. 2 (**a**), the actual (red dots) and estimated (blue line) IRF overlap (top-

right). On the top-left, a similar plotting is done for the ACF. The actual IRF is obtained after applying the LL method to the nonlinear hemodynamics approach, with a stimulus sequence defined by a unit pulse (a gaussian of around 200 mSecs of duration) at time 0. The actual ACF is calculated after feeding the hemodynamic approach with a gaussian white noise (see Fig. 3 (**b**) in Riera et al. 2004). The original BOLD signal (black) and the free-noise realization (blue) after fitting was performed are plotted (middle). The red line indicates the temporal course of the potential drift. The innovation process (which showed a histogram with gaussian distribution) and the block representation of the stimulation paradigm are shown in the bottom of the figure.

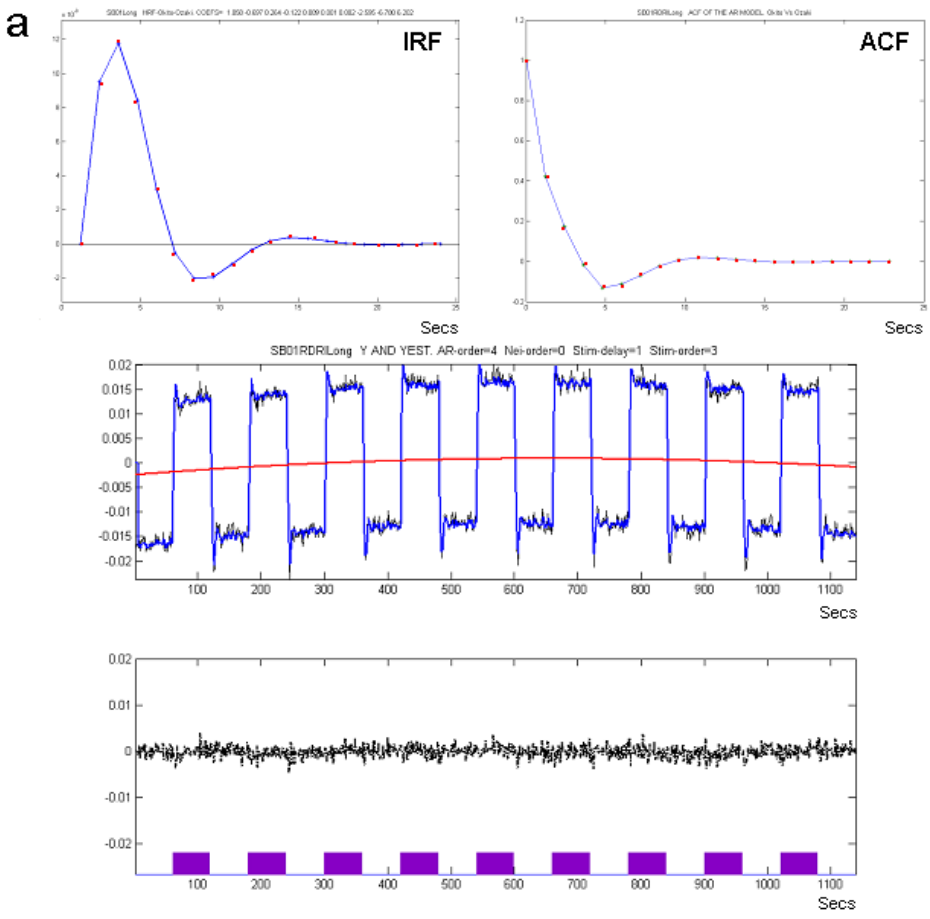

Fig. 2a. *The results obtained from fitting the ARx model to synthetic data created from the hemodynamics approach, with a transit time of 0.1 Secs, via the LL*

discretization method. The panel on the top shows the overlap of actual (red dots) and estimated (blue line) functions (IRF on the right and ACF on the left). In the middle, the original BOLD signal (black) and the free-noise realization (blue) after fitting overlap with the temporal course of the potential drift. The innovation process and the stimulation paradigm are shown in the bottom panel.

Fig. 2 (**b**) presents the values of the global parameters for which the AICc (top) reached a non-local minimum value. The vertical line identifies the minimum, while in the other two graphs the values of $p = 4$ and $q = 3$ can be read.

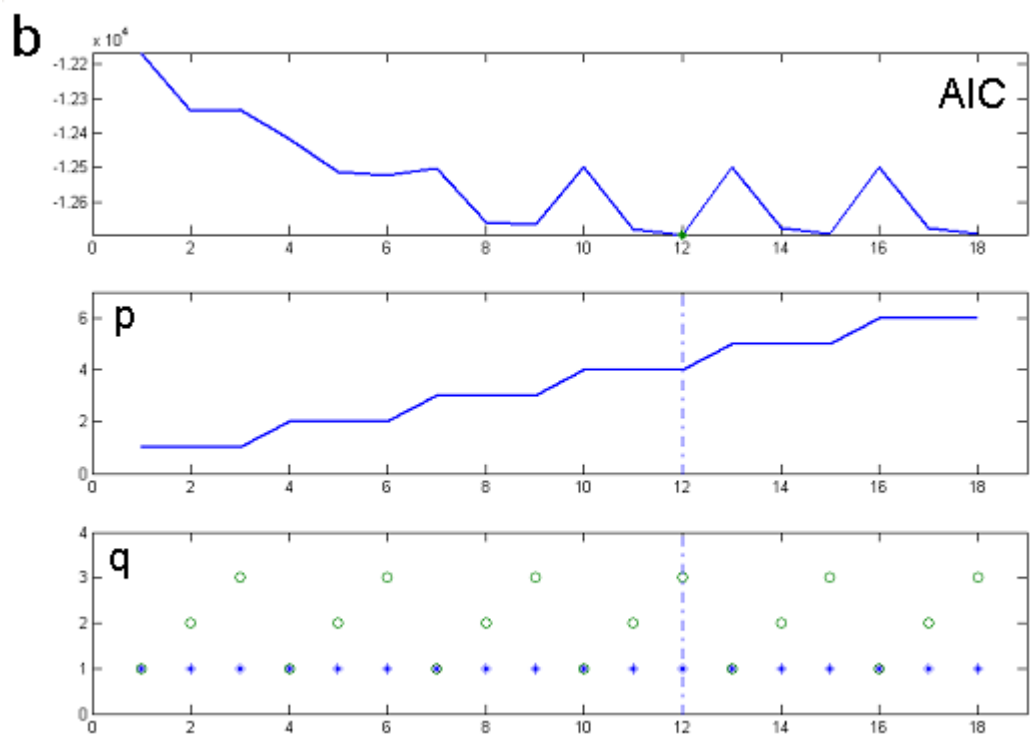

Fig. 2b. *The AICc is used for selecting a model. The global parameters for which the AICc (top) reached a non-local minimum value are exposed (i.e. vertical line).*

Motivated by previous results obtained by our group (see discussion in Riera et al. 2004), the authors were interested in assessing an extreme case where the IRF and ACF do not oscillate. This type of dynamics is observed in BOLD signals when the

transit time is increased considerably. In a recent paper, large transit times were reported using direct PET study with H(2)(15)O and (11)CO (Ito et al. 2003). Fig. 3 shows the same plots as Fig. 2 above, but in this case the transit time was set at 6 Secs. Both the IRF and the ACF exhibited an exponential decay behavior, which can be interpreted in the same way as a low-pass RC filter with small cutoff frequency (Fig. 3 (**a**)). In this example, the AICc reaches a minimum for $p = 5$ and $q = 3$. The histogram of the innovation process also showed a gaussian distribution in this case (Fig. 3 (**b**)).

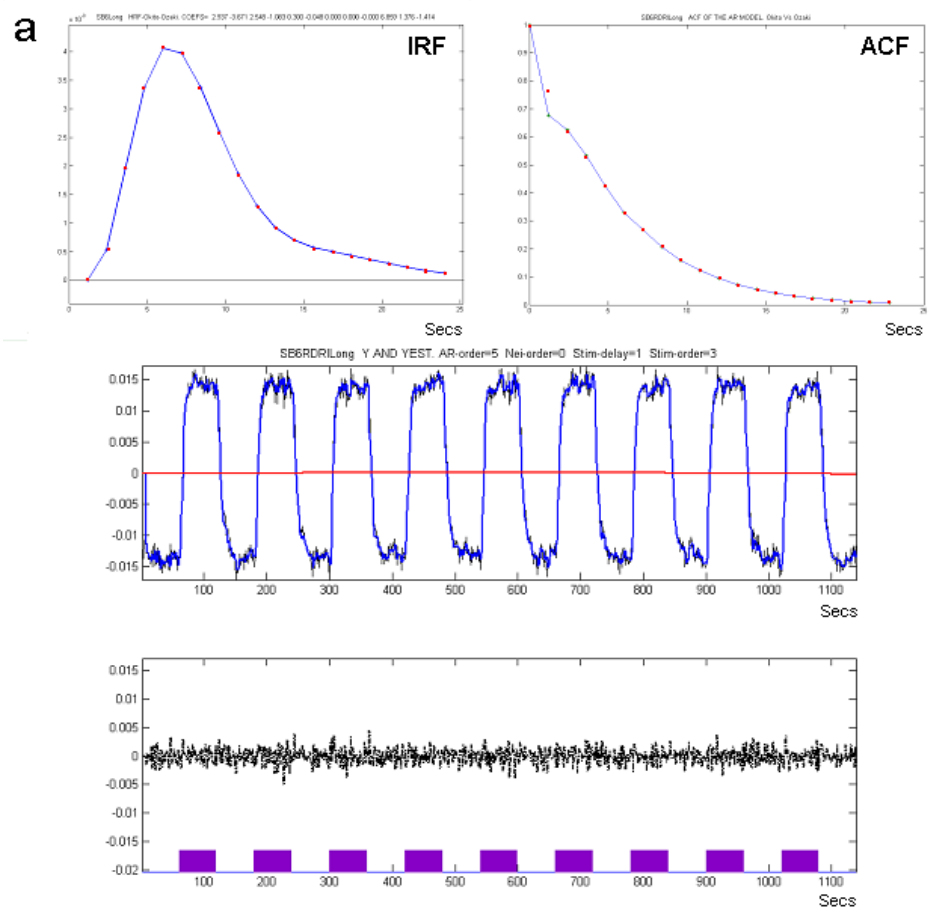

Fig. 3a. *The same panels as in Fig. 2 (**a**) are presented, but in this case a transit time of 6 Secs was used in the hemodynamics approach.*

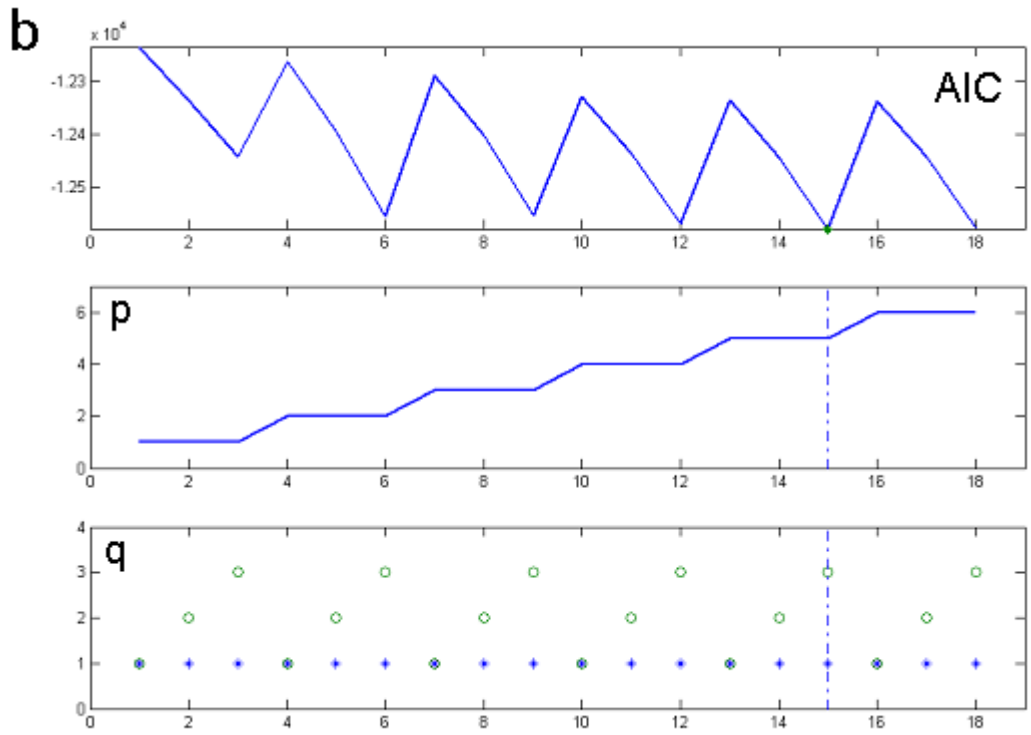

Fig. 3b. *The AICc is used for selecting a model (same panels as in Fig. 2 (**b**)).*

Furthermore, the authors consider it important to illustrate that the potential drift appearing in BOLD signals, which has been generally associated with artifacts, could be physiological in nature. In Fig. 4, a significant potential drift appears by simply increasing the strength of the randomness of the additive physiological noise (i.e. vector $\boldsymbol{g} = \{g_i\}$, equation (1) in Riera et al. 2004), which produces strong DC fluctuations in the dynamics (see red curve).

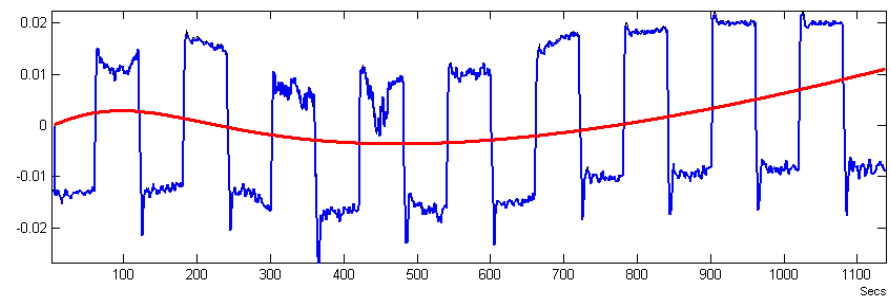

Fig. 4. *The potential drift originated from nonlinear fluctuations of the hemodynamics approach due to stochastic inputs.*

This was our chief motivation for including a polynomial series of "time" in the model (7) to account for the effect of potential drifts in a unified theoretical formalism.

Real data

The method was applied to actual BOLD data obtained under the two different experimental paradigms as described in methods section. The fMRI images were previously preprocessed as detailed in that section and with the help of the SPM99 toolbox. The θ-*MAP* figures will be presented only for the champion data in each experimental paradigm, but tables (with the hot-spots Talairach coordinates) and a figure (with their 3D representation on the *"Statistical Centroid"* of the McConnell Brain Imaging Centre, Montreal) will be used to summarize the results obtained in all subjects. Fig. 5 shows the results of the champion data for the case of the visual paradigm (checkerboard). In the top-left of the figure, the θ-*MAP* for different slices is presented, showing high precision in the localization of the V1 area. A damped oscillating IRF and ACF, respectively, are shown on the (top/middle)-right side. The superposition of the actual BOLD signal and the free-noise realization after model fitting for the hot-spot in the V1 area is plotted on the middle-left. The innovation process for that particular voxel (also with a histogram showing a gaussian distribution) and the specific experimental design paradigm are shown on the bottom-left. Finally, the AICc illustrates the model selectivity for dynamics complexity (bottom-right).

In order to establish credibility for the results reported by our method, an SPM99 analysis was performed after applying the classical SPM99 smoothing filter to the data. The on-off contrast T-test using "glass images" (maximum intensity projections) is presented in Fig. 6. The red mark shows the hot-spot in the visual primary area.

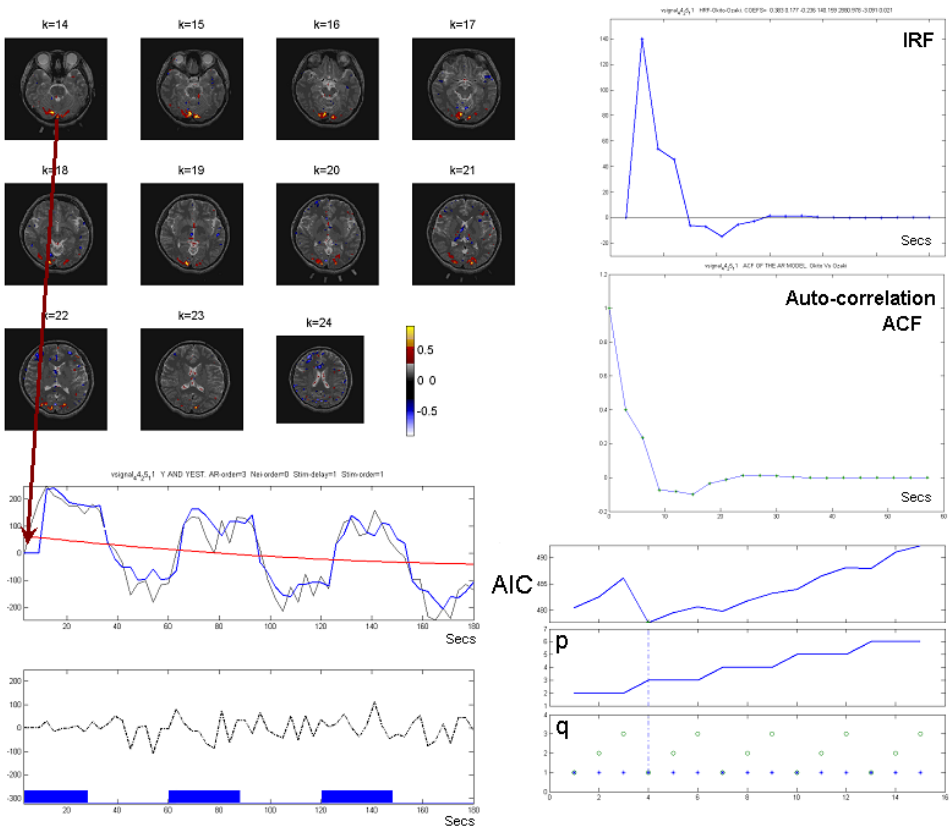

Fig. 5. *The results obtained from fitting the NN-ARx model to real data obtained by applying the visual paradigm (champion subject). The panels show the following: different slices of the θ-MAP superposed on the individual MRI (**top-left**), the IRF and ACF (**(top/middle)-right**), the superposition of the actual BOLD signal and the free-noise realization after model fitting for the hot-spot in V1 (**middle-left**), the innovation process for that particular voxel with the specific experimental design paradigm (**bottom-left**), and the AICc for model selection (**bottom-right**).*

Table-I summarizes the Talairach coordinates of hot-spots (obtained using both methods) for all subjects.

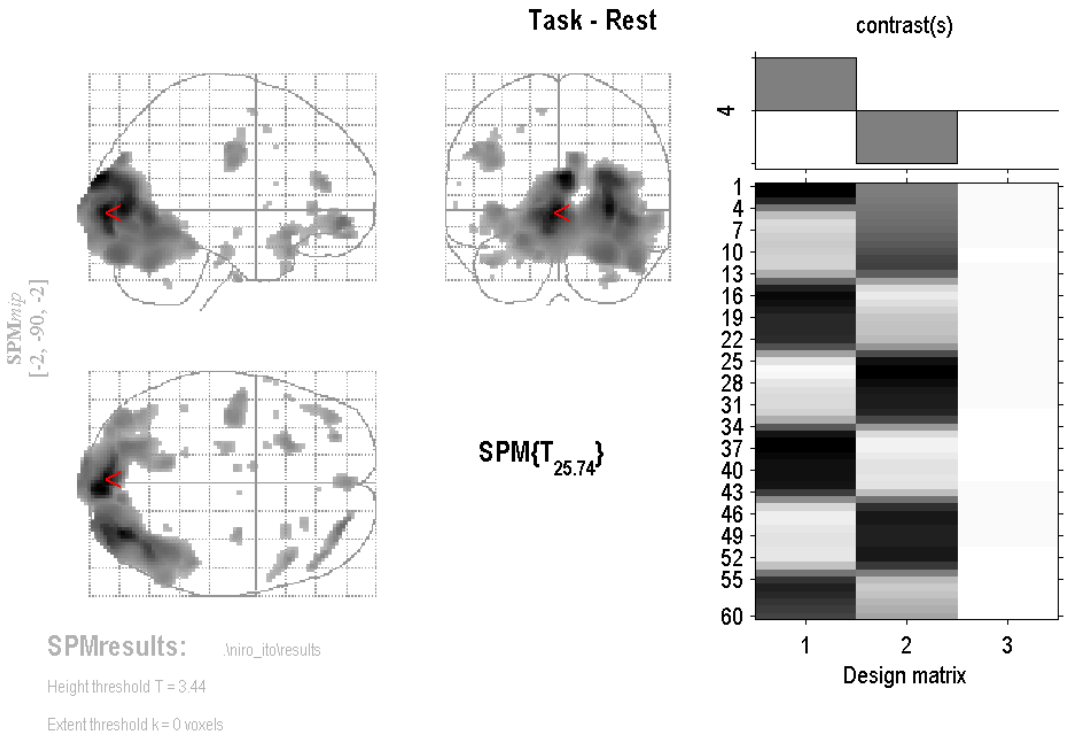

Fig. 6. *The output of the SPM99 Toolbox (i.e. T-test and design matrix for the visual paradigm).*

Subject	V1	
	SPM	**θ-*MAP***
#1	-2, -90, -2	12, -84, -4
#2	-6, -86, -6	-4, -90, -8
#3	4, -86, -4	14, -94, -2
#4	-8, -88, -14	14, -90, -2
#5	-2, -92, -14	10, -86, -14
#6	-10, -86, -4	10, -84, -10
#7	-12, -98, -4	12, -90, -8
#8	0, -82, -12	14, -84, -4
#9	-8, -92, -8	-4, -94, -6
#10	Error	10, -90, 8

Table-I. *The Talairach coordinates of hot-spots for the visual paradigm (V1) obtained via SPM analysis and the NN-ARx method.*

The results for the motor paradigm (right hand movement) are presented in Fig. 7. The θ-*MAP* displays activation in M1, Cerebellum and Cingulate Motor Area (CMA). Note that in this case, the IRF and the ACF for the M1 area exhibit an exponential

decay function-shape. Riera et al. (2004) reported the same result using the LL filter to estimate the parameters of the hemodynamics approach under a Kalman's system identification strategy.

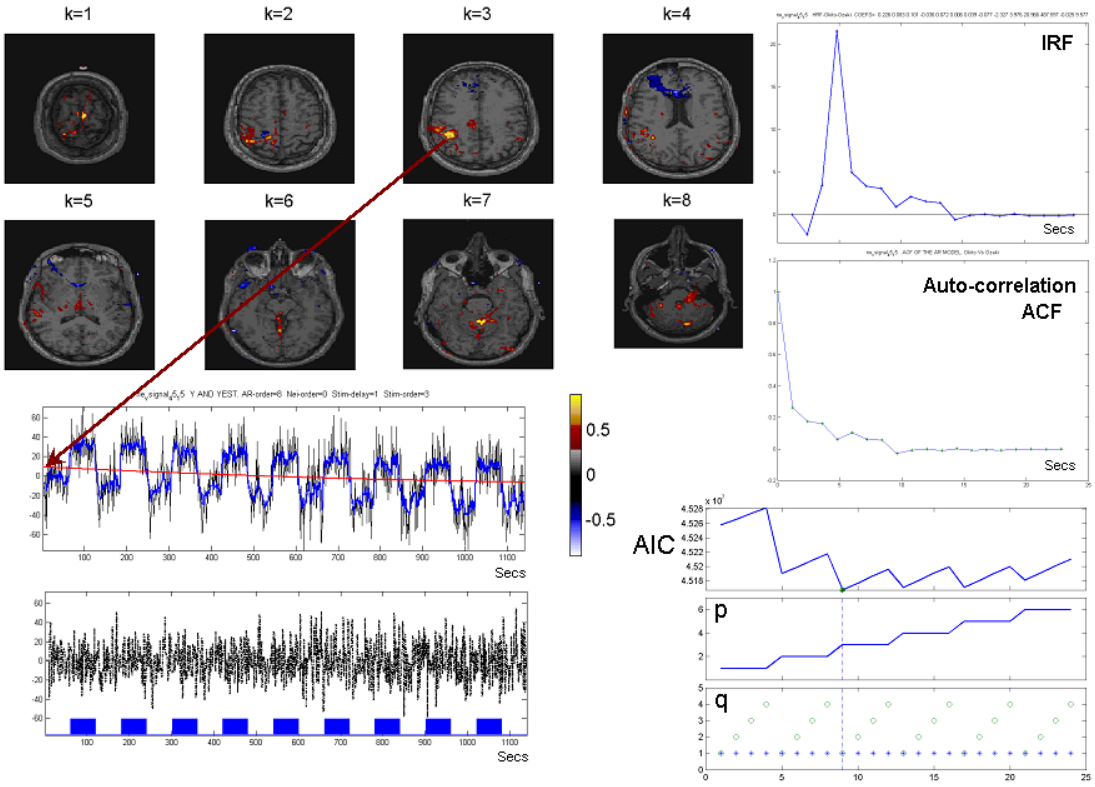

Fig. 7. *The results obtained from fitting the NN-ARx model to real data obtained by applying the motor paradigm (champion subject). These are the same panels as in Fig. 5, but in this case for the hot-spot in M1.*

In that paper, the estimated transit time was consistently larger than those stated in other comparative studies, which could explain the non-pronounced slope of the increasing and decreasing phases of the BOLD signal. Fig. 8 shows the T-test glass images for the same champion data. The red mark shows the hot-spot in the motor primary area. Table-II summarizes the results for all subjects in this particular experimental paradigm.

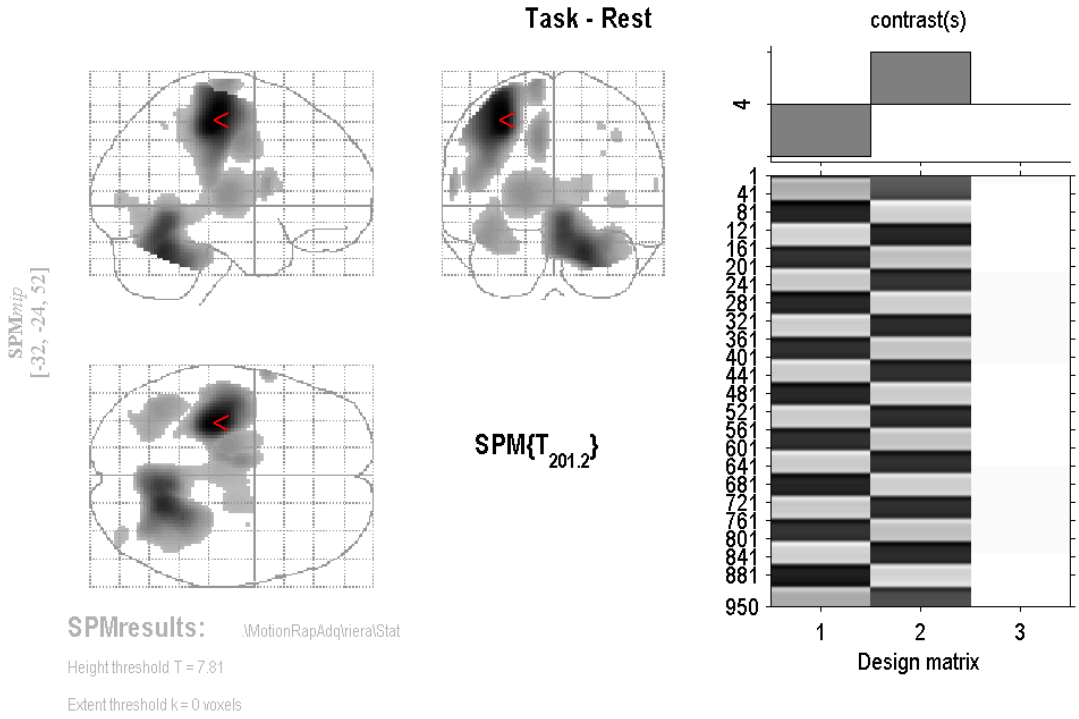

Fig. 8. *The output of the SPM99 Toolbox (i.e. T-test and design matrix for the motor paradigm).*

Subject	M1		Cerebellum		CMA	
	SPM	**θ-*MAP***	**SPM**	**θ-*MAP***	**SPM**	**θ-*MAP***
#1	-32, -24, 52	-28, -22, 54	18, -56, -28	18, -54, -26	-10, -8, 42	-6, -10, 46
#2	-40, -22, 64	-30, -22, 62	20, -60, 52	16, -64, -32	-4, -34, 50	-16, -20, 36
#3	-38, -20, 64	-38, -24, 52	18, -56, -26	18, -58, -26	-4, -6, 58	-6, -6, 44
#4	-38, -26, 56	-30, -26, 52	22, -56, -32	20, -60, -32	-6, -22, 42	-6, -4, 44
#5	-38, -14, 60	-32, -24, 54	18, -62, -24	20, -62, -24	-4, -10, 46	-4, -6, 44

Table II. *The Talairach coordinates of hot-spots for the motor paradigm (M1, CMA and Cerebellum) obtained via SPM analysis and the NN-ARx method.*

There is a very good correspondence between the significant active brain areas detected using T-test in the SPM99 toolbox and those reported using the θ-*MAP*, as proposed in our model. A 3D cluster representation of Tables I and II is shown in Fig. 9. The Statistical Centroid can be used to illustrate how those points (spheres) are grouped around V1, M1, Cerebellum and CMA areas for both, SPM T-test (red) and θ-*MAP* (green) method.

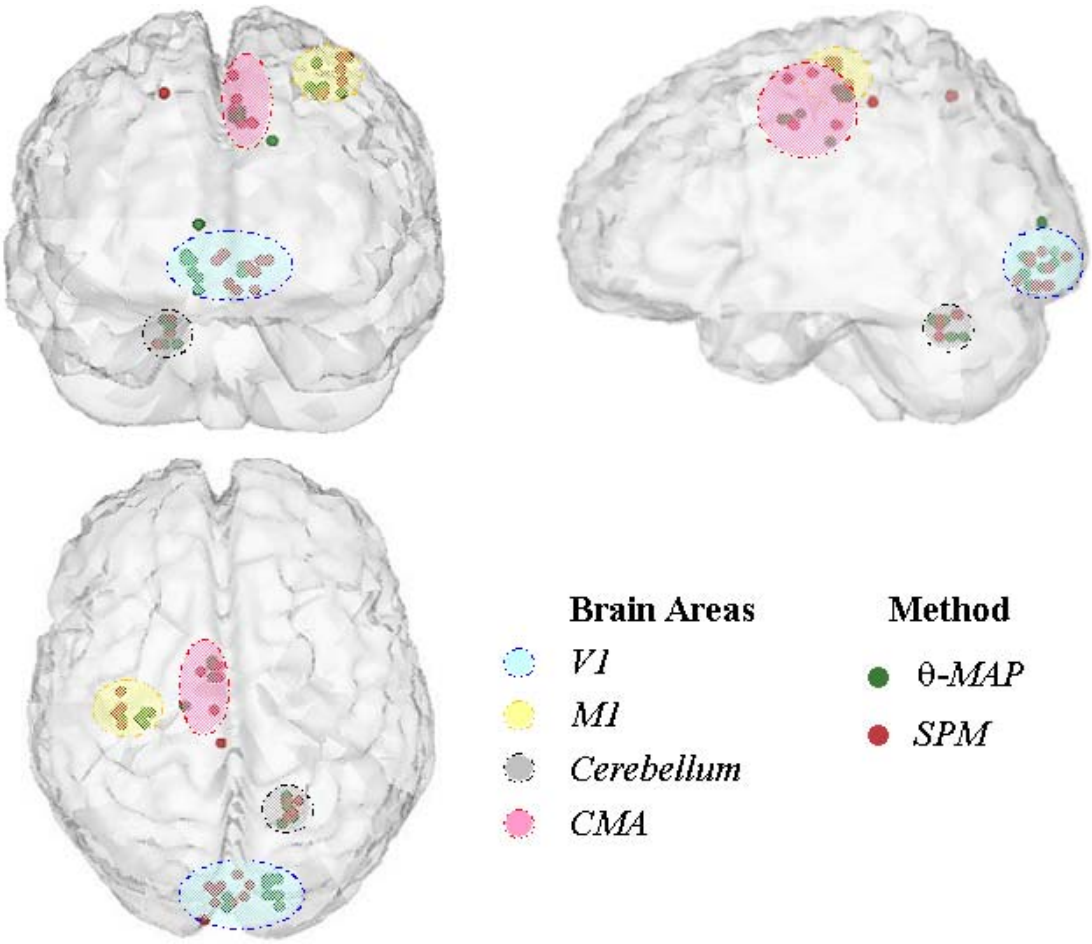

Fig. 9. *A cluster representation on the Statistical Centroid (Visual and Motor paradigms). Ellipsoids with different color transparencies limit the involved brain areas (V1, M1, CMA and Cerebellum). The hot-spots (spheres) obtained using the NN-ARx method and the SPM toolsbox are shown in green and red, respectively.*

Additionally, the contributions to each voxel from the near-neighborhood dynamics were also evaluated. The total influence $\pi_v = \sum_{v' \in \Omega_v} \left(\chi_v^{v'} \right)^2$ at v-th voxel could be interpreted as the magnitude of short-range vascular connections with its neighborhood. Fig. 10 shows a 3D representation of values π_v on a standardized anatomical image for the champion subject in the visual and motor task, respectively. It can be noted that there is no formation of particular spatial patterns, which could suggest a non-functional organization of the short-range connections at the level of

the vascular network. However, local correlations exhibited very strong anisotropic properties (see blue arrows in Fig. 10, where factors $\chi_v^{v'}$ were plotted in their respective directions), a fact which may be related to an inhomogeneous distribution of capillary beds in the cortex at a microscopic level (Harrison et al. 2002).

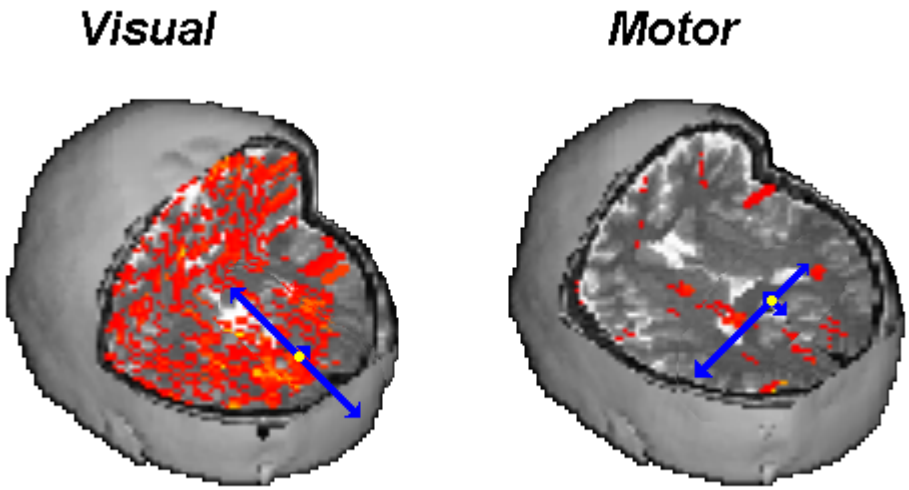

Fig. 10. *A 3D reconstruction of the spatial distribution of the total influence* π_v *for each paradigm is presented. The blue arrows show the 2D anisotropic factors (i.e. slice-view) for different directions in a particular voxel. The length of the arrow is proportional to the magnitude of the factor* $\chi_v^{v'}$.

Discussion

In this paper, the authors claim that θ-*MAP* can be used as an alternative way to perform fMRI analyses, with a clearer and more direct physiological interpretability of the results being achievable. When we do statistical modeling, there are two tactics: one of them is data-based with no theoretical model structure imposed, and the other is based on models with some degree of constraint coming from theoretical (i.e. physiological) assumptions. In modeling fMRI data we are interested in finding the IRF and also a noise structure model (i.e. ACF) in a common and unique formalism,

since these two magnitudes cannot be separated. It is clear from a time series analysis point of view that the noise structural model will be affected when particular assumptions about the IRF/HRF are made. The statistical method suggests that all you can rely on is the simultaneous fitness of the data (i.e. ML and AIC). In general, the fitness is better if constraints on both the IRF/HRF and the background noise are imposed, which explains why the authors have proposed the NN-ARx approach. Introducing a priori information via bayesian modeling can be considered adequate depending on whether the constraint (a forced assumption) is good or not. Past experience has shown the authors that these theoretical constraints are not always suitable, and this is mostly because the environment where the data is generated does not fit the idealistic theoretical models as much as we would like. The ACF of the fMRI data presented in the two experimental situations are good examples of this. The mechanism of generating fMRI data is much more complicated than using ARMA, but the point is to find the most useful approximation for "describing" (fitting) and "interpreting" the data. This discussion is motivated by studies carried out 30 to 50 years ago in mechanical, electrical and control engineering. Engineers at that time thought it was critical to carry out systems identification at the beginning. Therefore, a lot of experiments involving the inputting of periodic stimuli (i.e. sinusoidal waves) were performed during that period, with an analysis of the output data, showing that the best way to identify the system was to input white noise.

EEG and fMRI fusion

In this general formalism, it has become very clear how fMRI and EEG data fusion can be performed. It is the belief of the authors that, at the level of synapses, the electrophysiological activity coming from complementary brain areas starts affecting

the hemodynamical nonlinear response by the same hierarchical mechanisms as those activated during external stimulations via afferent pathways. Figure 11 illustrates the theoretical diagram that prompted us to propose a basic model for fMRI and EEG data fusion. In this case, a system of coupled oscillators is suggested, which differentiates two intrinsic dynamics at the v-th voxel: a fast state equation that accounts for synaptic activation α_t^v and a slow state equation that describes the vascular changes β_t^v.

$$\alpha_t^v = \sum_{v'=1}^{N_v}\sum_{k=1}^{l} A_k^{vv'}\alpha_{t-k}^{v'} + \sum_{k=0}^{r}\theta_k^v s_{t-k-d} + \varsigma_t^v \qquad \text{Fast dynamics}$$

$$\beta_t^v = \sum_{k=1}^{m} B_k^v \beta_{t-k}^v + \rho_v \sum_{k=0}^{s}\chi_k f\left(\alpha_{t-k}^v\right) + \zeta_t^v \qquad \text{Slow dynamics}$$

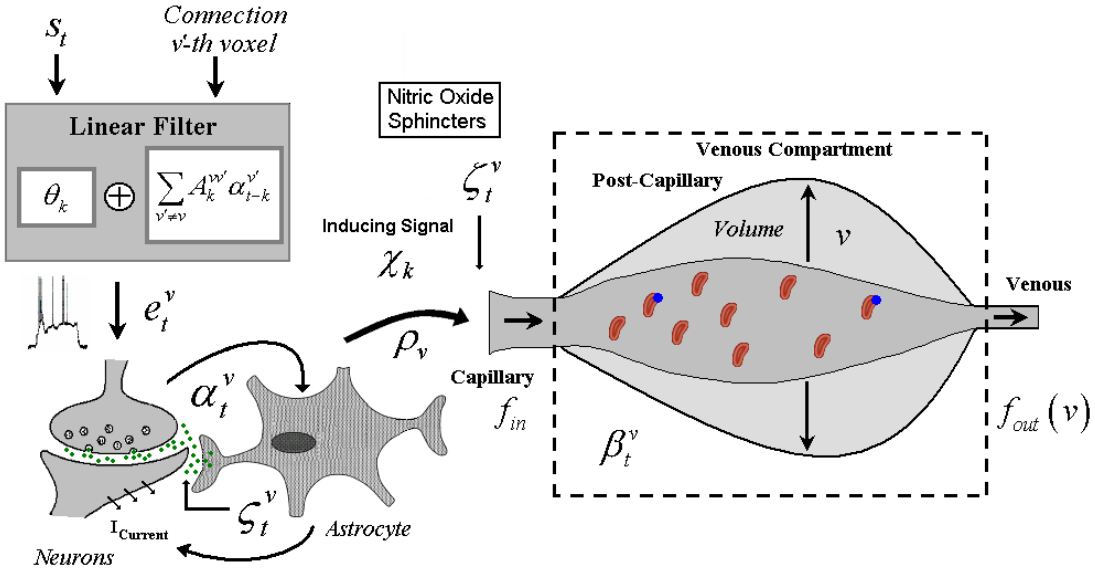

Fig. 11. *The diagram of the bottom-up model for the fusion of EEG and fMRI is illustrated. The system includes three blocks: a linear filter generating the evoked transients e_t^v, which plays the role of an integrator at the electrophysiological level inside the neurons, a fast dynamics linear subsystem emulating the neuron-astrocyte interrelationship at the synaptic level; and finally a slow dynamics linear subsystem (the dashed-line box) that mimics hemodynamics at the level of the micro-vascular*

building block. The connection between the fast and slow subsystems is only in one direction (i.e. synaptic activity creates a metabolics/oxygen demand, and will, therefore, induce an increase of blood flow via vascular regulation mechanisms). The factor ρ_v *could be an indicator of the susceptibility of the capillary bed to the flow-inducing signal. The magnitude* α_t^v *reflects changes in the synaptic activation, and may be associated with variations in the concentrations of specific neurotransmitters. The magnitude* β_t^v *will capture fluctuations in the blood volume v at the post-capillary venous compartment due to unbalanced inner* f_{in} *and outer* $f_{out}(v)$ *blood flows, which could also include the dependency of the BOLD signal with the concentration of de-oxy hemoglobin.*

Note that, a new evoked transient $e_t^v = \sum_{k=1}^{l} \sum_{v' \neq v} A_k^{vv'} \alpha_{t-k}^{v'} + \sum_{k=0}^{r} \theta_k^v s_{t-k-d}$ is defined at the neuronal-astrocyte scale, to which there are contributions not only from activations induced by the stimulus sequence but also from those emerging from interrelationships with other brain areas. The coefficients $A_k^{vv'}$ and B_k^v, associated with the fast and slow dynamical AR models, will determine the IRFs for two dissimilar dynamic modes. The random processes $\varsigma^v \sim N(0, \kappa_\alpha^v)$ and $\zeta^v \sim N(0, \kappa_\beta^v)$ define the system white noise introduced at the level of the synaptic cleft and the vasculature, respectively. The magnitudes α_t^v and β_t^v are hidden state-variables; hence, ML estimators cannot be explicitly determined from data/parameters, instead they can be calculated via recursive Kalman filter strategies (see Yamashita et al. 2004 and Riera et al. 2004 for the respective application to EEG and fMRI data). In the hemodynamics approach the function $f(\cdot)$ has been considered linear (Friston et al. 2000a). However, the authors believe that it is still enigmatic, so its definition could

be of great motivation for future cooperative works between physiologists and theoreticians. Valdes et al. (1999) have patented a very preliminary idea, but it uses the Bayesian formalism and lacks a model for the temporal dynamics of physiological processes.

The EEG observation equation is given by the solution of the forward problem for the particular volume conductor model (i.e. lead field vector $\vec{k}_{ev}$), with $\eta_t^e \sim N(0,\sigma_e)$ representing the instrumental error (i.e. white noise) and $\vec{m}^v$ being the orientation of brain electrical sources. That electrical source is a vector field that results from the spatial-temporal integration of thousands of small post-synaptic electric potentials in the micro-column range. By intuition, the magnitude α_t^v must be linearly proportional to the amplitude of such superimposed electric potentials. The time series of voltage differences between the electrode "e" and a common reference is symbolized by V_t^e.

$$V_t^e = \sum_{v=1}^{N_v} \vec{k}_{ev} \cdot \vec{m}^v \alpha_t^v + \eta_t^e \qquad e = 1,\cdots,N_e \quad \text{(Number of electrodes)}$$

The observation equation for the fMRI was deduced by Buxton et al. (1998), having established a direct nonlinear relationship between the BOLD signals and two intrinsic physiological variables: the blood volume and de-oxy hemoglobin concentration. However, a linear approach can be introduced to simplify that static non-linearity. The time series of BOLD signal is symbolized by y_t^v in our proposal, with $\eta_t^v \sim N(0,\sigma_v)$ representing instrumental error (i.e. white noise) and μ_t^v the above-mentioned potential drift. The effects produced by scaling factors in the BOLD signals are removed by using the voxel dependent parameter υ_v.

$$y_t^v = \mu_t^v + \upsilon_v \beta_t^v + \eta_t^v$$

The fact that temporal scales of EEG (mSecs) and fMRI (Secs) differ considerably, suggests the use of a linear operator χ_k to low-pass filter the fast synaptic activation would be appropriate. Note that equation (7) can be deduced from this more general model for data fusion. The model parameters $\Xi^v = \left\{ A_k^{vv'}, B_k^v, \rho_v, \upsilon_v, \kappa_\alpha^v, \kappa_\beta^v, \sigma_e, \sigma_v, \theta_k^v, \vec{m}^v, \gamma_k^v \right\}$ must be estimated for each voxel from EEG/fMRI data fusion. Some of these parameters can be inferred from other neuroimaging techniques (i.e. $A_k = \left\{ A_k^{vv'} \right\}$ matrices can be a priori designed from *diffusion tensor imagines* and structural MRI has been used to set $\vec{m}^v$ in the EEG inverse problem). The global parameters $\Lambda = \left(l, m, d, r, s, \delta \right)$ can also be estimated via AICc to determine model complexity.

Causality/Connectivity Patterns

As a whole, the methodology allowed us to estimate long-term connectivity by a simple examination of the unexplainable variances and covariances of the innovation process (however, this is just an exploratory method and a more exact formulation to include connectivity/causality interrelationships is now in the process of being revised). Fortunately, several works on this subject have recently appeared in the literature. Harrison et al. (2003), for example, have proposed a novel method which allows the estimation of nonlinear interactions between brain areas by using a multivariate AR with some extended bilinear variables. The model introduced a spatial variance-covariance matrix, and, by definition, the AR coefficients absorbed most of the correlation structure of the noise for the temporal varying BOLD signal. An alternative proposal to study nonlinear and non-synchronous interactions by testing linear and synchronous models against more general models was presented by

Lahaye et al. (2003). In that paper, it was proved that instantaneous interactions are less significant than interactions that consider the history of the BOLD data. They used time embedding and Volterra expansions to explore the interaction between two brain regions "*a*" and "*b*" (linear and nonlinear for both $a \to a$ and $b \to a$).

Non-linearities

Finally, the authors would like to discuss the recent tendency to incorporate nonlinear behaviors in the general linear models formalism. Friston et al. (1998b) have generalized the methodology that uses basis functions set to include nonlinear kernel contributions in the Volterra expansion. Josephs and Henson (1998) presented an excellent review that proposes linear approaches in the context of more general nonlinear dynamic models. In our opinion, a feasible generalization of the model proposed in this paper to include nonlinear BOLD dynamics can be carried out using state dependent AR coefficients in formula (7), such as in the case of the exponential AR model (Haggan and Ozaki 1981).

Acknowledgements

The authors would like to thanks Dr. Galka from "The Institute of Statistical Mathematics" in Tokyo for significant contributions to this paper during discussions. Dr. Aubert from Cuban Neuroscience Center in Havana has also supplied software for generating 3D illustrations. This study has been supported by Grant-in-Aid for Scientific Research (C) No.15500193, JSPS; JST/RISTEX, R&D promotion scheme for regional proposals promoted by TAO; and the 21st Century Center of Excellence (COE) Program (Ministry of Education, Culture, Sports, Science and Technology) entitled "A Strategic Research and Education Center for an Integrated Approach to Language, Brain and Cognition" (Tohoku University).

References

Berns GS, Song AW, Mao H (1999): Continuous functional magnetic resonance imaging reveals dynamic non-linearities of dose-response curves for finger opposition. The Journal of Neuroscience 19, RC17.

Birn RM, Saad ZS, Bandettini PA (2001): Spatial heterogeneity of nonlinear dynamics in the fMRI BOLD response. Neuroimage 14, 817-826.

Brockwell PJ, Davis RA (1987): Introduction to time series and forecasting. Springer Texts in Statistics.

Bullmore E, Brammer M, Williams SCR, Rabe-Hesketh S, Janot N, David A, Mellers J, Howard R, Sham Pak (1996): Statistical method of estimation and inference for functional MR image analysis. Magn. Reson. Med. 35, 261-277.

Burock MA, Dale AM (2000): Estimation and detection of event-related fMRI signals with temporally corrected noise: a statistically efficient and unbiased approach. Human Brain Mapping 11, 249-260.

Buxton RB, Wong EC, Frank LR (1998): Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn. Reson. Med. 39, 855-864.

Carew JD, Wahba G, Xie X, Nordheim EV, Meyerand E (2003): Optimal spline smoothing of fMRI time series by generalized cross-validation. Neuroimage 18, 950-961.

Dale AM (1999): Optimal experimental design for event-related fMRI. Human Brain Mapping 8, 109-114.

Friston KJ, Jezzard P, Turner R (1994): Analysis of functional MRI time series. Human Brain Mapping 1, 153-171.

Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R (1998a): Event-related fMRI: Characterizing differential responses. Neuroimage 7, 30-40.

Friston KJ, Josephs O, Rees G, Turner R (1998b): Nonlinear event-related responses in fMRI. Magn. Reson. Med. 39, 41-52.

Friston KJ, Mechelli A, Turner R, Price CJ (2000a): Nonlinear responses in fMRI: the balloon model, volterra kernels, and other hemodynamics. Neuroimage 12, 466-477.

Friston KJ, Josephs O, Zarahn E, Holmes AP, Rouquette S, Poline J-B (2000b): To smooth or not to smooth? Neuroimage 12, 196-208.

Friston KJ (2002): Bayesian estimation of dynamical systems: an application to fMRI. Neuroimage 16, 513-530.

Friston KJ, Penny W (2003): Posterior probability maps and SPMs. Neuroimage 19, 1240-1249.

Goutte C, Nielsen FA, Hansen LK (2000): Modeling the haemodynamic response in fMRI using smooth FIR Filters. IEEE Trans. Med. Imag. 19, 12, 1188-1201.

Haggan V, Ozaki T (1981): Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. Biometrika 68, 1, 189-196.

Harrison RV, Harel N, Panesar J, Mount RJ (2002): Blood capillary distribution correlates with hemodynamic-based functional imaging in cerebral cortex. Cerebral Cortex 12, 225-233.

Harrison L, Penny WD, Friston K (2003): Multivariate autoregressive modeling of fMRI time series. Neuroimage 19, 1477-1491.

Hopfinger JB, Buchel C, Holmes AP, Friston KJ (2000): A study of analysis parameters that influence the sensitivity of event-related fMRI analyses. Neuroimage 11, 326-333.

Huettel SA, McCarthy G (2000): Evidence for a refractory period in the hemodynamic response to visual stimuli as measured by MRI. NeuroImage 11, 547-553.

Huettel SA, McCarthy G (2001): Regional differences in the refractory period of the hemodynamic response: an event-related fMRI study. Neuroimage 14, 967-976.

Hurvich CM, Tsay C-L (1989): Regression and time series model selection in small samples. Biometrika 76, 2, 297-307.

Iadecola C (2002): Intrinsic signals and functional brain mapping: caution, blood vessel at work. Cerebral Cortex 12, 223-224, CC Commentary.

Ito H, Kanno I, Takahashi K, Ibaraki M, Miura S. (2003): Regional distribution of human cerebral vascular mean transit time measured by positron emission tomography. Neuroimage 19, 1163-1169.

Jimenez JC, Ozaki T (2003): Local linearization filters for non-linear continuous-discrete state space models with multiplicative noise. Int. J. Control 76, 12, 1159-1170.

Josephs O, Turner R, Friston K (1997): Event-related fMRI. Human Brain Mapping 5, 243-248.

Josephs O, Henson RNA (1998): Event-related functional magnetic resonance imagein: modeling, inference and optimization. Phil. Trans. R. Soc. Lond. B 354, 1215-1228.

Katanoda K, Matsuda Y, Sugishita M (2002): A spatio-temporal regression model for the analysis of functional MRI data. Neuroimage 17, 1415-1428.

Kruggel F, von Cramon DY (1999): Modeling the hemodynamic response in single-trial functional MRI experiments. Magn. Reson. Med. 42, 787-797.

Lahaye P-J, Poline J-B, Flandin G, Dodel S, Garnero L (2003): Functional connectivity: studying nonlinear delayed interactions between BOLD signals. Neuroimage 20, 962-974.

Lange N, Zeger SL (1997): Nonlinear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging. J. R. Stat. Soc. Appl. Stat. 46, 1-29.

Locascio JJ, Jennings PJ, Moore CI, Corkin S (1997): Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance brain imaging. Human Brain Mapping 5, 168-193.

Marchini JL, Smith SM (2003): On bias in the estimation of autocorrelations for fMRI voxel time-series analysis. Neuroimage 18, 83-90.

Magistretti PJ, Pellerin L (1999): Cellular mechanisms of brain energy metabolism and their relevance to functional brain imaging. Phil. Trans. R. Soc. Lond. B 354, 1155-1163.

Mandeville JB, Marota JJA, Ayata C, Zaharchuk G, Moskowitz, MA, Rosen BR, Weisskoff RM (1999): Evidence of cerebrovascular postarteriole windkessel with delayed compliance. J. Cereb. Blood Flow Metab. 19, 6, 679-689.

Marrelec G, Benali H, Ciuciu P, Pelegrini-Issac M, Poline J-B (2003): Robust bayesian estimation of the hemodynamic response function in event-related BOLD fMRI using basic physiological information. Human Brain Mapping 19, 1-17.

Purdon PL, Solo V, Weisskoff RM, Brown EN (2001): Locally regularized spatiotemporal modeling and model comparison for fMRI. Neuroimage 14, 912-923.

Rajapakse JC, Kruggel F, von Cramon DY (1998): Modeling hemodynamic response for analysis of functional MRI time-series. Human Brain Mapping 6, 283-300.

Riera JJ, Watanabe J, Kazuki I, Naoki M, Aubert E. Ozaki T, Kawashima R. (2004): A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signals. Neuroimage 21, 547-567.

Valdés P, Riera J, Bosch J, Aubert E, Virue T, Morales F, Trujillo N, Fuentes ME, Soler J (1999): System and Methods for the tomography of the primary electric current of the brain and the heart. ASSIGNEES: Centro de Neurociencias de Cuba, Cuba (OCPI, (11) Certify No. 22550). Patent Cooperation Treaty (PCT) WO 99/10816.

Woolrich MW, Ripley BD, Brady M, Smith SM (2001): Temporal autocorrelation in univariate linear modeling of fMRI data. Neuroimage 14, 1370-1386

Worsley KJ, Liao CH, Aston J, Petre V, Duncan GH, Morales F, Evans A (2002): A general statistical analysis for fMRI data. Neuroimage 15, 1-15.

Worsley KJ, Friston K (1995): Analysis of fMRI time-series revisited-again. Neuroimage 2, 173-181.

Worsley KJ, Poline J-B, Friston KJ, Evans AC (1997): Characterizing the response of PET and fMRI data using multivariate linear models. Neuroimage 6, 305-319.

Weisskoff RM, Baker J, Belliveau J, Davis TL, Kwong KK, Cohen MS, Rosen BR (1993): Power spectrum analysis of functionally-weighted MR data: what's in the noise? [Abstract] Proc. Soc. Magn. Reson. Med. 1, 7.

Yamashita O, Galka A, Ozaki T, Biscay R, Valdes PA (2004): Recursive penalized least squares solution for dynamical inverse problems of EEG generation. Human Brain Mapping 21, 221-235.

Appendix I. Laplacian **pre-filtering**

The most commonly used filtering technique applied to raw fMRI data is the gaussian kernels. Volumetric smoothing kernels with a full-width half-max in the range of 10 mm have been suggested as optimal for event-related fMRI experiments (Hopfinger et al. 2000). The use of such filters introduces an undesirable spatial correlation. In our model, we should be able to suitably distinguish between the two main sources of correlation, both physiological and nuisance contributions.

At first, it is assumed that $\boldsymbol{x}_t = \left(x_t^1, \cdots, x_t^{N_v}\right)^t$ is a random field with a variance-covariance matrix $Cov\left(\boldsymbol{x}_t, \boldsymbol{x}_t\right) = \Sigma_{\mathbf{xx}}$. Therefore, a spatial whiteness is obtained by applying the inverse filter on the original data $\boldsymbol{y}_t = \Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x}_t$. After some exploratory searching, we found evidence of the local nature of that instantaneous correlation, which suggested the use of the Laplace inverse operator filter $\Sigma_{\mathbf{xx}} = L^{-1}$ to eliminate it and to guarantee instantaneous noise orthogonality. It can be mathematically formulated as $y_t^v = 6x_t^v - \sum_{v' \in \Omega_v} x_t^{v'}$. The symbol Ω_v represents the neighborhood of the v-th voxel.

Appendix II. ***ARMAx*** model: *generalized impulse response and autocorrelation* functions

In this appendix, for mathematical simplicity, superscript to label voxels has not been included. The AutoRegressive Moving-Average (ARMA) model is defined by:

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \cdots + \psi_q \varepsilon_{t-q} \qquad \text{(II-1)}$$

where ε_t is a temporally independent random element (called "*innovation*") with zero mean and finite variance σ^2. Equation (II-1) can be rewritten using the backshift (or lag) operator B ($Bx_t = x_{t-1}$) as follows:

$$(1 - \phi_1 B - \cdots - \phi_p B^p) y_t = (1 + \psi_1 B + \cdots + \psi_q B^q) \varepsilon_t$$

or using algebraic polynomial notation:

$$\Phi(B) y_t = \Psi(B) \varepsilon_t \qquad \text{(II-2)}$$

An $ARMA(p,q)$ model is "*stationary*" if all the zeros of $\Phi(B)$ lie outside the unit circle (some authors refer to this fact as the "*causality*" condition (Brockwell and Davis 1987). Under this condition, an $ARMA(p,q)$ model has "*Wold representation*"

$$y_t = (1 + h_1 B + h_2 B^2 \cdots) \varepsilon_t = H(B) \varepsilon_t, \qquad \text{(II-3)}$$

that is, the process y_t is represented by the infinite sum of the past innovations $\varepsilon_t, \varepsilon_{t-1}, \cdots$. The series of the coefficients h_k is called the *impulse response function* since the function $h(k)$ can be regarded as the response at times k to a unit pulse input at time 0 (Fig. II (**a**)).

Substituting equation (II-3) into equation (II-2), the following equation holds:

$$\Phi(B)H(B)\varepsilon_t = \Psi(B)\varepsilon_t \qquad \text{(II-4)}$$

Because both sides of (II-4) must be identical as a polynomial of B, equating the coefficients of every degree B^k leads to the system of equation:

$$\begin{aligned} -\phi_1 + h_1 &= \psi_1 \\ -\phi_2 - \phi_1 h_1 + h_2 &= \psi_2 \\ -\phi_3 - \phi_2 h_1 - \phi_1 h_2 + h_3 &= \psi_3 \\ &\vdots \end{aligned} \qquad \text{(II-5)}$$

The impulse response function h_k can be obtained from the coefficients ϕ_k and ψ_k by solving system (II-5) recursively from the topmost equation. In practical applications, it is usually sufficient to obtain h_k up to some large value k. The system of equations (II-5) can be summarized by:

$$h_k - \sum_{k'=1}^{\min(k,p)} \phi_{k'} h_{k-k'} = \psi_k \qquad \text{(II-6)}$$

with $h_0 = 1$ and $\psi_k = 0$ for $k > q$. It should be noted that once the Wold representation is obtained, the *autocorrelation* function $R(\tau)$ of output y_t could be easily computed from the auto-covariance function $C(\tau)$.

$$C(\tau) = E(y_t y_{t+\tau}) = \sigma^2 \sum_{k=0}^{\infty} h_k h_{k+\tau}$$

$$R(\tau) = C(\tau) / C(0)$$

$$R(\tau) = \sum_{k=0}^{\infty} h_k h_{k+\tau} \Big/ \sum_{k=0}^{\infty} h_k h_k \qquad \text{(II-7)}$$

where $E(\cdot)$ denotes expected value. The Fig. II (**b**) illustrates how the white noise ε_t is colored at the output of the ARMA model. A symbolic function $R(\tau)$ is also plotted.

Henceforth, the extension to an ARMA model with an exogenous variable (ARMAx) is immediate. An $ARMAx(p,q,r)$ model is defined by:

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \cdots + \psi_q \varepsilon_{t-q} + \\ + \theta_0 s_t + \theta_1 s_{t-1} + \cdots + \theta_r s_{t-r} \qquad \text{(II-8)}$$

where s_t is an (deterministic) exogenous variable independent from the innovation process ε_t. Using the backshift operator, equation (II-8) is rewritten as:

$$\Phi(B) y_t = \Psi(B)\varepsilon_t + \Theta(B) s_t \qquad \text{(II-9)}$$

Note that $\Theta(B) = \sum_{k=0}^{r} \theta_k B^k$. Then, the equation (II-9) can be finally rewritten as:

$$\Phi(B)\, y_t = \overline{\Psi}(B)\overline{\varepsilon}_t \qquad \text{(II-10)}$$

The extended variable $\overline{\varepsilon}_t = (\varepsilon_t, s_t)^t$ is a vector that comprises both the innovation process and the exogenous variable. The extended polynomial $\overline{\Psi}(B) = (\Psi(B) \quad \Theta(B))$ is a vector which summarizes two equivalent moving-average polynomials for process ε_t and variables s_t. In the same way as for a pure ARMA model, if all the zeros of $\Phi(B)$ lie outside the unit circle, the process y_t has a Wold representation. Then, the equation (II-9) can be finally rewritten as:

$$y_t = \overline{H}(B)\overline{\varepsilon}_t \qquad \text{(II-11)}$$

where $\overline{H}(B) = [H(B) \quad \Gamma(B)]$ represents the generalized impulse response function, which includes the original impulse response function h_k (i.e. obtained also from II-6) involved in the calculation of autocorrelation function and a new impulse response function g_k for the exogenous variable. It is not difficult to show that the coefficients g_k related to the polynomial $\Gamma(B)$ can be obtained by recursively calculating:

$$g_k - \sum_{k'=1}^{\min(k,p)} \phi_{k'} g_{k-k'} = \theta_k \text{ , with } g_0 = \theta_0 \qquad \text{(II-12)}$$

Fig. II (**c**) exemplifies the function $g(k)$, and is interpreted as the response at times k to an exogenous unit pulse input at time 0.

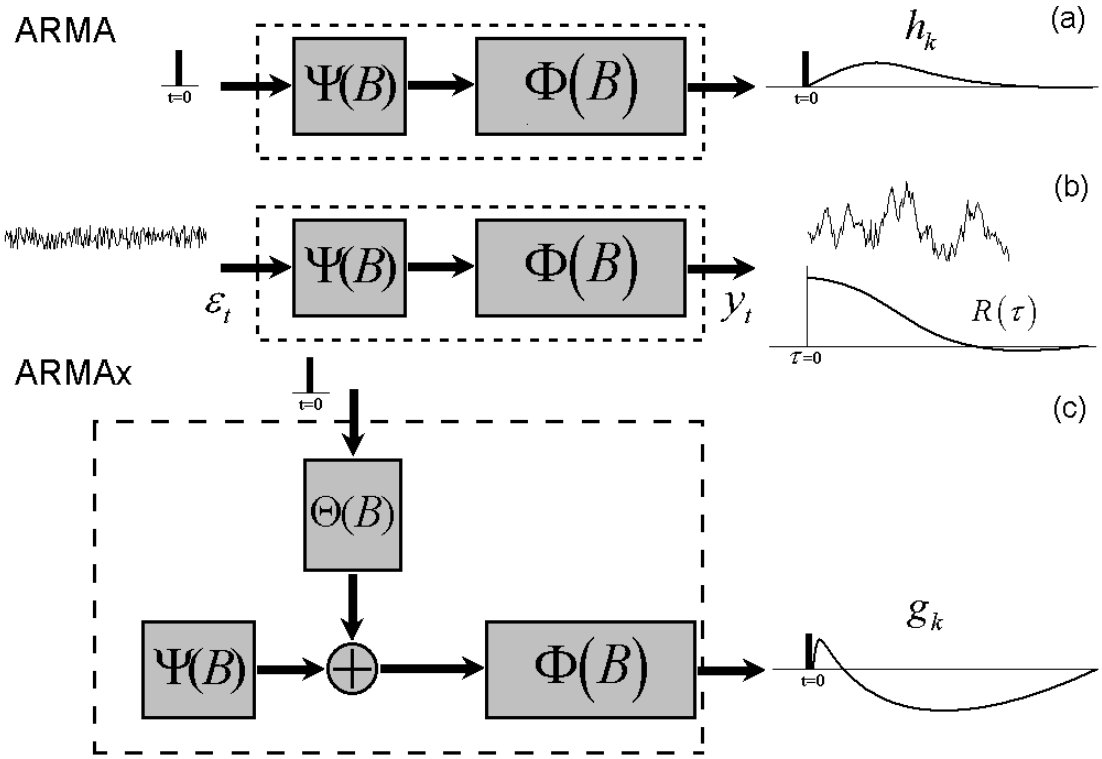

Fig. II. *The IRFs for an ARMA model: (**a**) without (h_k) and (**c**) with (g_k) exogenous variables are illustrated. Panel (**b**) shows a symbolic picture, where a white noise input is colored by the linear system, producing a random process with an ACF $R(\tau)$ at the output.*

Appendix III. *ML estimators and AICc for model selection*

The system identification contains two sequential steps: a) the estimation of model parameters Ξ^v for the NN-ARx in each voxel, and b) a model selection (i.e. determine the global parameters Λ) by using the AIC. The likelihood function for the time series $y_1^v,\cdots,y_N^v$ of BOLD signal in the voxel *v* is given by:

$$p(y_1^v,\cdots,y_p^v,y_{p+1}^v,\cdots,y_N^v;\Xi^v)=p(y_1^v,\cdots,y_p^v)\prod_{t=p+1}^{N}p(y_t^v\mid y_{t-1}^v,\cdots,y_{t-p}^v;\Xi^v)$$

Where $p(y_1^v,\cdots,y_p^v)$ is the distribution for the initial value $y_1^v,\cdots,y_p^v$ and

$$p(y_t^v\mid y_{t-1}^v,\cdots,y_{t-p}^v;\Xi^v)=\frac{1}{\sqrt{2\pi\sigma_v^2}}\exp\left\{-\frac{1}{2\sigma_v^2}\left(y_t^v-\mu_t^v-\sum_{k=1}^{p}\phi_k^v y_{t-k}^v-\boldsymbol{X}^v\boldsymbol{\xi}_{t-\Delta}^v-\sum_{k=0}^{r}\theta_k^v s_{t-k-d}\right)\right\}^2$$

The time series on the nearest neighbor voxels are approximately handled as the exogenous variables (i.e. not random variables). Under the assumption of the gaussian innovation, the log-likelihood function can be represented as:

$$\ell\left(\Xi^v\right)=-\frac{1}{2}\log 2\pi\sigma_v^2-\frac{1}{2\sigma_v^2}\sum_{t=p+1}^{N}\left(y_t^v-\mu_t^v-\sum_{k=1}^{p}\phi_k^v y_{t-k}^v-\boldsymbol{X}^v\boldsymbol{\xi}_{t-\Delta}^v-\sum_{k=0}^{r}\theta_k^v s_{t-k-d}\right)^2 \qquad \text{(III-1)}$$
$$+\log p(y_1^v,\cdots,y_p^v)$$

The third term in the formula (III-1) can be neglected since the first and second terms are dominant when the number of data N increases. Therefore, the ML estimators of $\hat{\Xi}^v$ can be obtained by maximization of the sum of the first and the second terms in (III-1), which actually corresponds to the simple least squares estimators.

The global parameters Λ should be determined in an objective way using any of the information criteria (i.e. such as AIC, BIC, SIC). In this paper, we would like to employ the corrected version of AIC, (i.e. AICc), which has been reported to improve the correctness of the AR order selection in small sample simulations (Hurvich and Tsay 1989). The AICc for our model is given by:

$$AICc(\Lambda) = \sum_v \left[-2\ell\left(\hat{\Xi}^v\right) + \frac{2N(N_p+1)}{N-N_p-2} \right] = N\sum_v \log\hat{\sigma}_v^2 + \frac{2N(N_p+1)}{N-N_p-2} N_v \quad \text{(III-2)}$$

Where N_p is the number of parameters Ξ^v in the model and N_v is the number of voxels. We determine Λ by minimization of AICc (III-2). It was implicitly assumed that voxels were statistically independent.