ROBUST BRAIN ACTIVATION DETECTION IN FUNCTIONAL MRI

David Afonso, João Sanches *

Institute for Systems and Robotics
Instituto Superior Técnico
1049-001 Lisbon, Portugal

Martin H. Lauterbach (MD)

Faculty of Medicine
University of Lisbon
1649-028 Lisbon, Portugal

ABSTRACT

Functional *Magnetic Resonance Imaging* (MRI) is today one of the most important non-invasive tools to study the brain from a functional point of view. The blood-oxygenation-level-dependent (BOLD) signal is used to detect the activated regions based on the assumption that in these regions the metabolic activity increases. The normal procedure is the application of known sequences of stimulus and find out the brain regions whose activation sequence is correlated with the applied stimulus.

This inference problem is difficult because the BOLD signal is very week and noisy. The underlying information is embedded in a large number of other signal related with the normal brain activity and in the noise introduced by the MRI scanner. Furthermore, the *hemodynamic impulse response function* (HRF), needed to know the expected BOLD response to a given stimulus, is usually unknown and is not constant across the whole brain.

In this paper a robust Bayesian algorithm is proposed to detect regions where the activation patterns are correlated with the applied stimulus. The activation process is modeled by using binary explicative variables and the HRF is estimated at each location according to a physiological model proposed by the authors in [1].

Monte Carlo tests using synthetic data are performed to evaluate the performance of the algorithm and results with real data are compared with the ones obtained by a neurologist with the commercial package *BrainVoyager*.

Index Terms— Functional MRI, Bayesian, Estimation.

1. INTRODUCTION

Functional Magnetic Resonance Imaging (fMRI) is an emergent and powerful technique used in several clinical scopes. One of the most important applications is in the *brain imaging* field to detect brain regions involved in particular tasks. This modality is based on the assumption that in the activated regions the metabolic activity increases, leading to a change on the amount and oxygenation of the blood on that region. This change may be indirectly measured from the *blood-oxygenation-level-dependent* (BOLD) signal, that measures the endogenous magnetic contrast(ratio) between oxyhaemoglobin (diamagnetic) and deoxyhaemoglobin (paramagnetic). Hence, increased blood volume reduces the local concentration of deoxygenated hemoglobin causing an increase in the magnetic resonance (MR) signal on a T2 or T2*-weighted image [2].

During the acquisition a set of MRI volumes are acquired at constant time intervals, synchronized with the stimulus signal. The goal is to locate the regions where the respective *times courses* (1D signal associated with the temporal evolution of each voxel intensity) present temporal activation patterns correlated with the applied stimulus, called *paradigm* in the fMRI scope.

This task is usually difficult because the BOLD signal has a low *signal-to-noise-ratio* (SNR), is corrupted by noise introduced by the MRI scanner, the images may be misaligned due motions occured during the acquisition and the underlying information to be estimated is embeded in a large number of other signal related with the normal brain activity. Furthermore, the *hemodynamic impulse response function* (HRF), needed to correctly deconvolve the BOLD signal, is usually unknown and is not constant across the whole brain [3]. The most common algorithms described in the literature to detect the activated regions is based on the general linear model (GLM). The general approach is to express the observed response variable in terms of a linear combination of explanatory variables (EVs) [4], and make use of classical statistics (T or F tests) to infer activity, using e.g. a p-value threshold where the HRF is first estimated and used for all voxels. The HRF may be modeled in a pure heuristic basis or in a physiological basis. In this last approach the some of the underlying physiological processes involved in the BOLD signal generation are modeled, e.g. the Balloon Model [5] which is often used and augmented [6]. These last models are complex, with several parameters and difficult to tune. In this paper we make use of the linear, *infinite impulse response* (IIR) *physiologically based hemodynamic* (PBH) model proposed by the authors, that presents an appealing trade-off between complexity and accuracy as shown in [1].

In this paper an algorithm is proposed where the EV's and the HRF are jointly estimated in a Bayesian framework using the *maximum a posteriori* (MAP) criterion in a local basis. This means that the HRF is not assumed constant across the whole brain. Only two assumptions are performed about the HRF to be estimated at each location: *i)* it is smooth and *ii)* it is an admissible response of the PBH model for an unknown set of parameters that must be estimated.

Monte Carlo tests are presented for synthetic data and the error probabilities obtained with the proposed method are compared with the standard classical procedures used in several software, e.g. the Statistical Parametric Mapping (SPM), FMRIB Software Library and BrainVoyager. Results using real data are compared with the ones obtained by an experimented neurologist with the commercial software package *BrainVoyager*.

The results obtained with real data are consistent with the ones obtained by the medical doctor and, as it will be shown later, the proposed method manages to detect regions that are not detected by the SPM-GLM method.

The rest of the paper is organized as follows. Section 2 formulates the problem and describes the detection algorithm. Section 3 presents the experimental results and Section 4 concludes the paper.

*Correspondent author: João Sanches (jmrs@ist.utl.pt). This work was supported by Fundao para a Cincia e a Tecnologia (ISR/IST plurianual funding) through the POS Conhecimento Program which includes FEDER funds.

2. PROBLEM FORMULATION AND DETECTION

The data is formed by a set of L volumes acquired at constant time intervals. The evolution of the signal BOLD, associated with a single voxel, along the time, is called *time course* and in this paper the time courses are processed independently. The following observation model is used,

$$y(n) = h(n) * \sum_{k=1}^{N} \beta_k p_k(n) + \eta(n) \quad (1)$$

where $\eta(n)$ is *additive white Gaussian noise* (AWGN), $h(n)$ is the HRF of the brain tissues, $p_k(n)$ are the stimulus signals along time and β_k are unknown binary variables to model the activation of the voxel by the k[th] stimulus.

Each voxel, after the application of a given paradigm may be activated by one or more applied stimulus ($\exists_k : \beta_k = 1$) or may not be activated at all ($\forall_k : \beta_k = 0$). Therefore, each *time course* is the response of $H(z)$ to the following input signal, $x(n) = \sum_k \beta_k p_k(n)$ where the binary EV's β_k must be estimated.

In this paper we describe a Bayesian *Statistical Parametric Mapping* algorithm (SPM) based on the *maximum a posteriori* (MAP) criterion called *SPM-MAP*. The proposed algorithm jointly estimates the vector $\mathbf{b} = \{\beta_1, \beta_2, ..., \beta_N\}^T$, associated with each voxel and the corresponding hemodynamic response, $h(n)$, which can be denoted in vectorial form, $\mathbf{h} = \{h(1), h(2), ..., h(N)\}^T$.

The hemodynamic signal is assumed to be the response of an IIR *linear time invariant system* proposed in [1] and described by the following third order transfer function

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3}} \quad (2)$$

where the coefficients b_k and a_k must be estimated.

The estimation process is performed by minimizing an energy function depending on the binary unknowns β_k, on the hemodynamic response $h(n)$ and on the observations $y(n)$. The direct coefficient estimation of the $H(z)$ is a difficult task because it is not easy to define simple priors for these coefficients based on the desired time response $h(n)$. Therefore, to overcome this difficulty, instead of estimating the a_k and b_k coefficients, a L length FIR is estimated, $\mathbf{g} = \{g(1), g(2), ..., g(L)\}^T$. In each iteration this estimated response is *projected* into the $H(z)$ space, i.e., a set of coefficients a_k and b_k are estimated in order to minimize $\|g(n) - h(n)\|$ by estimating the coefficients a_k and b_k. The PBH a_k and b_k estimated coefficients are then used to compute a new estimation of $\mathbf{h}$, $h^{t+1}(n) = g(n)$ for $1 \leq n \leq L$, which is used to obtain a new estimate of the binary unknowns β_k^{t+1}.

The *maximum a posteriori* (MAP) estimation is obtained by minimizing the following energy function

$$E(y, \mathbf{x}(\mathbf{b}), \mathbf{h}) = E_y(\mathbf{y}, \mathbf{x}(\mathbf{b}), \mathbf{h}) + E_b(\mathbf{b}) + E_h(\mathbf{x}(\mathbf{b})) \quad (3)$$

where $\mathbf{x}(\mathbf{b})$ is the input signal depending on $\mathbf{b}$, the *data fidelity term* is $E_{\mathbf{y}}(\mathbf{y}, \mathbf{x}(\mathbf{b}), \mathbf{h}) = -\log p(\mathbf{y}|\mathbf{x}(\mathbf{b}))$ and the prior terms associated to the unknowns to be estimated, $\mathbf{b} = \{\beta_1, ...\beta_N\}$ and $\mathbf{h} = \{h(1), ..., h(L)\}$ are $E_{\mathbf{b}}(\mathbf{b}) = -\log(p(\mathbf{b}))$ and $E_{\mathbf{h}}(\mathbf{x}(\mathbf{b})) = -\log(p(\mathbf{h}))$ respectively. These priors incorporate the *a priori* knowledge about the unknowns to be estimated: *i)* β_k are binary and *ii)* $h(n)$ is smooth.

The estimation process is performed in the following three steps,

$$\mathbf{b}^t = \arg\min_{\beta} E(\mathbf{y}, \mathbf{x}(\mathbf{b}^{t-1}), \mathbf{h}^{t-1}) \quad (4)$$

$$\mathbf{g} = \arg\min_{h} E(\mathbf{y}, \mathbf{x}(\mathbf{b}^t), \mathbf{h}^{t-1}) \quad (5)$$

$$\mathbf{h}^t = \text{Proj}_{FIR}\left[\text{Proj}_{IIR}(\mathbf{g})\right] \quad (6)$$

where $()^t$ means estimation at t^{th} iteration and $Proj$ stands for the projection operation by using the *minimum square error* (MSE) criterion and implemented with the *Shanks* algorithm [7].

Assuming independence of the observations and adopting the *additive white Gaussian noise* (AWGN) model, $p(\mathbf{y}|\mathbf{x}(\mathbf{b}), \mathbf{h}) = \prod_{i=1}^{L} p(y(i)|(x * h)(i))$ where $p(y(i)|(x * h)(i)) = \mathcal{N}((x * h)(i), {\sigma_y}^2)$ and ${\sigma_y}^2$ is the variance of the noise. The parameters β_k to be estimated are also assumed independent, which means, $p(\mathbf{b}) = \prod_{i=1}^{N} p(\beta_k)$ where $p(\beta_k)$ is a bi-modal distribution defined as a sum of two Gaussian distributions centered at zero and one, with variance ${\sigma_\beta}^2$, $p(\beta_k) = \frac{1}{2}\left[N(0, \sigma_\beta^2) + N(1, \sigma_\beta^2)\right]$ because β_k are binary variables, $\beta_k \in \{0, 1\}$. In order to better approximate the binary answer, the σ_β parameter should be as small as possible but numerical stability reasons prevent the adoption of too small values. The prior term $E_b(b)$ may therefore be written as

$$E_b(\mathbf{b}) = \sum_{k=1}^{N} \left[\frac{2\beta_k^2 - 2\beta_k + 1}{4\sigma_\beta^2} - \log\left(\cosh\left[\frac{2\beta_k - 1}{4\sigma_\beta^2} \right] \right) \right].$$

To impose smoothness on the estimated hemodynamic response, $h(n)$ is assumed to be a *Markov Random Field* (MRF), which means, that $p(\mathbf{h})$ is a *Gibbs* distribution, $p(\mathbf{h}) = \frac{1}{Z_h} e^{-\alpha \sum_{n=2}^{N} (h(n) - h(n-1))^2}$ where α is a parameter that tunes the smoothing degree for $h(n)$ and Z_h is a partition function. The energy function (3) is therefore

$$\begin{aligned} E(\mathbf{y}, \mathbf{x}, \mathbf{h}) = & \frac{1}{2\sigma_y^2} \sum_{n=1}^{L} \left[y(n) - \sum_{k=1}^{N} \beta_k \left(h * p_k\right)(n) \right]^2 + \\ & \sum_{k=1}^{N} \left[\frac{2\beta_k^2 - 2\beta_k + 1}{4\sigma_\beta^2} - \log\cosh\left[\frac{2\beta_k - 1}{4\sigma_\beta^2} \right] \right] + \\ & \alpha \sum_{n=2}^{L} \left(h(n) - h(n-1)\right)^2 \end{aligned} \quad (7)$$

The minimization of the equation (7) is performed in three steps

$$\nabla_{\mathbf{b}} E(\mathbf{y}, \mathbf{x}(\mathbf{b}), \mathbf{h}^t) = 0 \rightarrow \mathbf{b}^{t+1} \quad (8)$$

$$\nabla_{\mathbf{h}} E(\mathbf{y}, \mathbf{x}(\mathbf{b}^{t+1}), \mathbf{h}) = 0 \rightarrow \mathbf{g} \quad (9)$$

$$h^{t+1} = \text{Proj}_{FIR}\left[\text{Proj}_{IIR}(\mathbf{g})\right] \quad (10)$$

where $\nabla_{\mathbf{b}}$ and $\nabla_{\mathbf{h}}$ are the gradient vectors of $E(\mathbf{y}, \mathbf{x}(\mathbf{b}), \mathbf{h})$ with respect to $\mathbf{b}$ and $\mathbf{h}$ respectively. Proj_{FIR} denotes the extraction of the first L samples of the infinite response $h(n)$ and Proj_{IIR} the estimation of the parameters a_k and b_k of $H(z)$ from the finite response $\mathbf{g}$. These three steps are repeated until convergence is achieved. To accomplish the desired binary nature of $\hat{\mathbf{b}}$, at the end, the following threshold is applied to $\hat{\beta}_k$

$$\hat{b}_k = \begin{cases} 0 & \hat{\beta}_k < 0.5 \\ 1 & otherwise. \end{cases} \quad (11)$$

and this is the final activation estimation that provides information on whether the brain area represented in the corresponding voxel was activated by each of the paradigm stimulus or not.

3. EXPERIMENTAL RESULTS

In this section, results using synthetic and real data are presented to illustrate the application of the algorithm.

3.1. Monte Carlo tests with synthetic data

In this section Monte Carlo tests are presented in order to evaluate the performance of the algorithm. Two synthetic binary images of 128x128 pixels where generated, which represent a single BOLD slice signal, as can be seen overlapped in Fig. 1-a). In it, colored voxels (red, yellow and white) where activated by, at least, a stimulus paradigm and the black pixels where not activated at all. So according to the mathematical notation presented above, red: $\mathbf{b} = \{1, 0\}^T$; yellow: $\mathbf{b} = \{0, 1\}^T$; white: $\mathbf{b} = \{1, 1\}^T$ and black: $\mathbf{b} = \{0, 0\}^T$, which is the activation ground truth to be estimated for each voxel.

(a) (b) (c)

Fig. 1. (a) Synthetic image representing a single BOLD slice with brain areas activated by two paradigms (red and yellow), with a functional overlapping region (white) and non activated areas (black). *SPM-MAP* activation detection results for (b) $\sigma_y = 0.5$ and (c) $\sigma_y = 1$.

The BOLD signal, $y(n)$, is generated by using the model described by the equation (1). A reasonable two stimuli block-design paradigm, $p_1(n)$ and $p_2(n)$, of 10 seconds task duration followed by a 30 second rest period each in 5 epochs, were used in order to obtain a non superposition of $p_1(n)$ and $p_2(n)$ while allowing for the BOLD signal to decay to rest. The true impulse HRF signal, $h(n)$, was generated from a representative IIR, selected from the PBH estimation on real single-event data [1], and the following noise energies were used: $\sigma_y = \{0.2, 0.5, 0.7, 0.8, 1\}$.

This generated synthetic data is equivalent to $2 \times 128 \times 128 = 32768$ independent $y(n)$ time-courses, containing all possible combinations for the $\mathbf{b}$ vector. These are used on Monte Carlo tests to compute the P_e. The results obtained are graphically presented in Fig. 1 and in Table 1. These values were computed as the ratio of the total number of wrong estimations over the total number of Monte Carlo tests (32768).

σ_y	0.2	0.5	0.7	0.8	1
$P_e(\%)$	0.0427	0.0916	0.168	0.260	4.27
$\dot{P}_e(\%)$	0	0	0.0244	0.0245	1.51
$\ddot{P}_e(\%)$	0	0	0.0073	0.0061	0.513

Table 1. Monte Carlo P_e of *SPM-MAP* for several values of σ_y. Spatial correlation correction is exemplified in $\dot{P}_e$ and $\ddot{P}_e$ where one and two isolated pixels were dismissed, respectively.

It is important to point-out that realistic σ_y noise values would be situated between 0.2 and 0.5, for the data used. In this range the method achieves values of $P_e < 0.1\%$. Furthermore, for the very high noise amount of $\sigma_y = 1$ the P_e stays below 5%, resulting in the bottom image in Fig. 1 (notice that when looking at Fig. 1, the intuitive notion on the error probability might seem higher because the two images are overlapped).

The accuracy of the method can be improved if spacial correlation information is included, removing several of those isolated, spatially uncorrelated, voxels. For illustration purposes, the P_e is recalculated after removing areas of one ($\dot{P}_e$) and two ($\ddot{P}_e$) isolated voxels in an 8 voxels neighborhood. The resultant error probabilities (see Table 1) decreases for all the noise amounts, yielding null for the 0.2 and 0.5 σ_y values. The HRF average estimations are presented in Fig. 2 for $\sigma_y = \{0.05, 0.5, 1\}$. Other methods, more complex, may be used to model the spatial correlation in order to reduce even more the error probability, such as, *Total Variation* (TV) edge preserving denoising filters.

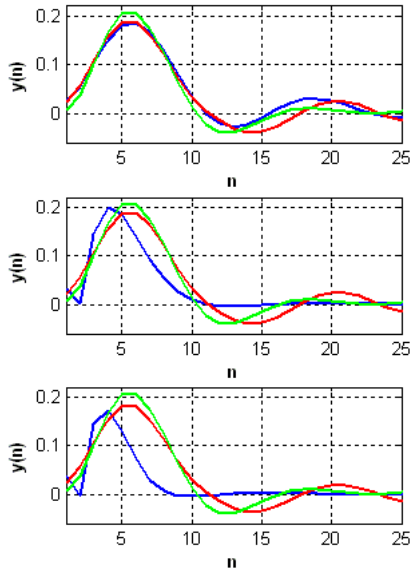


Fig. 2. Mean HRF estimation results considering for $\sigma_y = \{0.05, 0.5, 1\}$ from top to down. The Real HRF used for data generation is in green, the estimated FIR average in red, and the estimated IIR average in blue.

3.2. Real data

In this section results using real data are presented an compared with the ones obtained by the medical doctor by using the *Brain Voyager* software. Two volunteers with no history of neurological or psychiatric diseases participated on motor and trajectory generation block-designed paradigms during fMRI data acquisition on a *Philips Intera Achieva Quasar Dual 3T* whole-body system with a 8 channel head-coil. T2*-weighted echo-planar images (EPI), 23cm square field of view with a 128×128 matrix size resulting in an in-plane resolution of $1.8 \times 1.8mm$ for each $4mm$ slice, echo time = $33ms$, flip angle = 20^o were acquired with a $TR = 3000ms$.

The fMRI data was preprocessed with the standard procedures implemented in the *Brain Voyager* software, namely decrease of data distortions due to motion or other phase changes over time (registration) and spacial smoothing. This data was then statistically processed by the *Brain Voyager SPM-GLM* and *SPM-MAP* algorithms, and the results are plotted in Fig. 3. Since the obtainable brain maps by *SPM-GLM* highly depends on the selected p-values a neurologist provided the results (third column), for each data set, which he considers more correct (*reference* result). Since this result is subjective, he also provided two other results which he considers *loose* (second column) and *restricted* (fourth column). The first column displays the *SPM-MAP* results.

Visual inspection of the results in Fig. 3 show some expected resemblance between the neurologist *reference* solution, obtained with the *SPM-GLM* algorithm, and the brain maps obtained by the proposed *SPM-MAP* algorithm. The small differences may be explained

by the structural different approaches used by both methods about the HRF. In the *SPM-GLM* this response is considered space invariant while in the *SPM-MAP* methods it is jointly estimated in each time course.

In several of the brain map image sets there are brain regions detected as activated by *SPM-MAP* that were not detected by *SPM-GLM* as shown in the (g) section of Fig.3. These are unlikely false positives. Considering that the error probability has been shown considerably low (see section 3.1), the probability of several false positives occuring grouped in a small image area, instead of randomly dispersed in the image, is very small. Furthermore, as shown in the (g) section of Fig. 3 the *time courses* associated with these areas present a clear correlation with the paradigm which allow the conclusion they are true activated areas.

4. CONCLUSIONS

In this paper the SPM-MAP algorithm for detection of brain activated regions in the scope of *Functional MRI* is described. The algorithm is design in a Bayesian framework using the *maximum a posteriori* (MAP) criterion and estimates simultaneously the activation binary variables and the *hemodynamic response function* (HRF) at each voxel location. The performance of the algorithm was evaluated by using synthetic data and a Monte Carlo methodology where used to compute the average error probability, P_e, for several amount of Gaussian noise.

Results using real data were compared with the ones provided by a neurologist obtained with the traditional SPM-GLM method. The comparison has shown similar results without need of any parameter defined by the medical doctor that induces an undesirable subjectivity in the results. Additionally, it was shown the ability of the proposed algorithm to detect areas that are not detect by the SPM-GLM method.

5. REFERENCES

[1] David Miguel Afonso, Joo Sanches, and Martin Hagen Lauterbach, "Neural physiological modeling towards a hemodynamic response function for fMRI," Aug 2007, pp. 1615–1618.

[2] S. Ogawa, R. S. Menon, D. W. Tank, S. G. Kim, H. Merkle, J. M. Ellermann, 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*, vol. 64, no. 3, pp. 803–812, Mar 1993, Comparative Study.

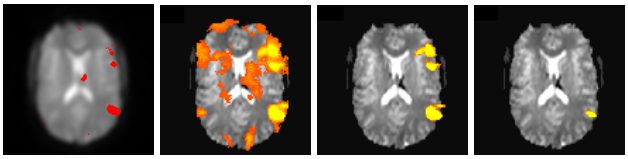
[3] Daniel A. Handwerker, John M. Ollinger, and Mark D'Esposito, "Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses," *Neuroimage*, vol. 21, no. 4, pp. 1639–1651, Apr 2004.

[4] K. J. Friston, "Analyzing brain images: Principles and overview," in *Human Brain Function*, R.S.J. Frackowiak and K.J. Friston and C. Frith and R. Dolan and J.C. Mazziotta, Ed., pp. 25–41. Academic Press USA, 1997.

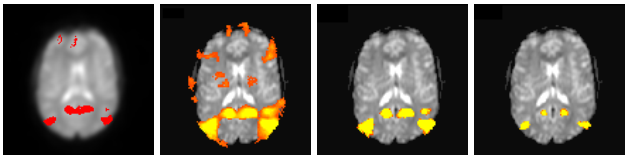
[5] R. B. Buxton, E. C. Wong, and L. R. Frank, "Dynamics of blood flow and oxygenation changes during brain activation: the balloon model," *Magn Reson Med*, vol. 39, no. 6, pp. 855–864, Jun 1998.

[6] K.J. Friston, A. Mechelli, R. Turner, and C.J. Price, "Nonlinear responses in fMRI: The Balloon model, Volterra kernels and other hemodynamics," *NeuroImage*, vol. 12, pp. 466–477, 2000.

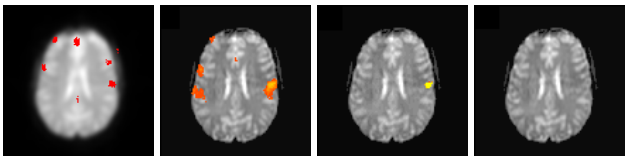
[7] John L. Shanks, "Recursion filters for digital processing," *Geophysics*, vol. 32, pp. 33–51, 1967.



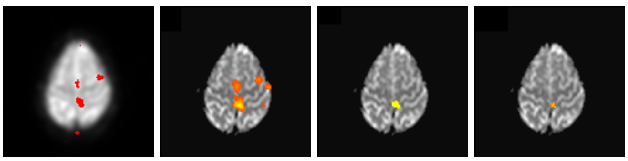
(a) Verb generation paradigm results.



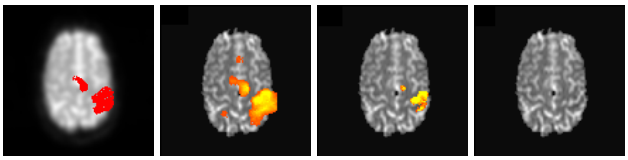
(b) Trajectory generation paradigm results.



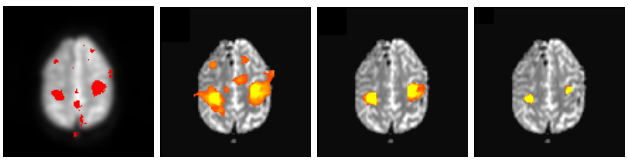
(c) Right foot motor paradigm results.



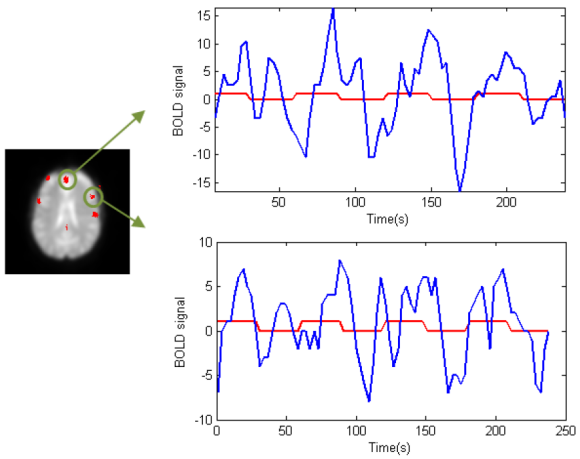
(d) Right foot motor paradigm results.



(e) Right hand motor paradigm results.



(f) Bilateral hands motor paradigm results.



(g) Non detect areas with the SPM-GLM method.

Fig. 3. *SPM-MAP* activity detection results (left) on real data compared against the ones obtained by a neurologist with the *SPM-GLM* method ($3rd$ column - *reference*, $2nd$ column - *restricted* and $4th$ column - *loose*).