Estimating the delay of the fMRI response

C.H. Liao[1], K.J. Worsley[1,2], J-B. Poline[3],
J.A.D. Aston[1,2], G.H. Duncan[4], A.C. Evans[2]

[1]*Department of Mathematics and Statistics*
[2]*Montreal Neurological Institute, McGill University, Canada*
[3]*Service Hospitalier Frédéric Joliot, CEA, 91401 Orsay, France*
[4] *Centre de Recherche en Sciences Neurologiques, Université de Montréal*

May 30, 2005

Keywords: fMRI, hemodynamic response function, delay, linear regression, non-linear regression, bias reduction.

Corresponding Address:
Chuanhong Liao
Department of Mathematics & Statistics
McGill University
805 Sherbrooke Street West
Montréal, Québec, Canada, H3A 2K6
Phone: (514)398-3800, Fax: (514)398-3899
E-mail: liao@math.mcgill.ca

Abstract

We propose a fast, efficient, general, simple, valid and robust method of estimating and making inference about the delay of the fMRI response modeled as a temporal shift of the hemodynamic response function (HRF). We estimate the shift unbiasedly using two optimally chosen basis functions for a spectrum of time shifted HRFs. This is done at every voxel, to create an image of estimated delays and their standard deviations. This can be used to compare delays for the same stimulus at different voxels, or for different stimuli at the same voxel. Our method is compared to other alternatives, and validated on an fMRI data set from an experiment in pain perception.

1 Introduction

Brain electrical activity is not directly measured; instead the human hemodynamic responses to brief periods of neural activity are delayed and dispersed in time. This is modeled by convolving the stimulus with the hemodynamic response function (HRF) of the vascular system of the brain. A second component of the delay may come from delays in processing the stimulus or reacting to it. Thus fMRI measures the subsequent demand for oxygenated blood that follows about 6 seconds after the neuronal response. It is this delay or latency of the fMRI response of about 6 seconds that we wish to estimate in this paper.

It is important to estimate precisely this delay since it can inform on the neural dynamics of brain activity. Indeed, even if the metabolic filter may vary from one brain region to another, leading to some difficulties in interpreting differences in delay between regions, the hemodynamics can generally be assumed to remain constant in time for a given localisation. This allows us to interpret differences in delay within the same region. It has to be noted that some authors have shown variation in delay across brain regions with plausible interpretations, for instance in Kruggel *et al.* (1999, 2000) in a language task.

Several methods have been proposed for estimating parameters of the HRF. The first, proposed by Lange and Zeger (1997), models the HRF by a gamma function, then estimates the delay and dispersion parameters by non-linear regression (see e.g. Seber and Wild, 1989). This requires time consuming iterative methods, such as the Gauss-Newton method, which can be unstable when the algorithm does not converge. Purdon *et al.* (2001) also used non-linear regression for a model of the HRF based on physiological considerations. This method is also time consuming to implement. An intriguing alternative, proposed by Rajapakse *et al.* (1998), uses a non-iterative method that fits a Gaussian function to the HRF. This method relies on a special property of the Fourier transform of the Gaussian function: the phase is linearly related to the delay, and the log modulus is linearly related to the squared dispersion. However the Gaussian is not a good choice for the HRF, since it implies negative delays and its overall shape does not match the well-known post-peak dip in the HRF (Friston *et al.*, 1998). Other methods of estimating the delay have been based on the phase of the response to a periodic stimulus (e.g. Thierry *et al*, 1999), but this does not work for non-periodic stimuli. The method of Saad (2001) is more direct: using the Hilbert transform, the cross-correlation between the stimulus and the fMRI data is searched for the maximum. Again this may be time consuming if different stimuli have different delays, and there is no theoretical estimate of the standard error of the delay, so that this must be estimated empirically from repeated runs.

In Friston *et al.* (1998) latency differentials were modeled using a linear model with a first order Taylor expansion of evoked responses with respect to latency or delay, so that the model includes the expected response and the derivative with time of this response. This is a simple and relatively robust approach that enabled the authors to make inferences about latency differences between one trial type and another in the order of a few hundred milliseconds. The shortcoming of the approach described in Friston *et al* (1998) is that the parameter scaling the contribution of the temporal derivative term depends upon the magnitude of the evoked response and not just the latency itself. In fact the approach adopted in Friston *et al* (1998) is more properly understood in terms of a second order Taylor expansion around a response with zero magnitude and delay where the temporal derivative term represents the

second order interaction between amplitude and delay. In other words, the method is exact only if there is no delay and zero magnitude.

To overcome these shortcomings we have explored the use of estimation and inference procedures using a linear model that focuses on the delay itself. We do this by using two optimally chosen basis functions of the hemodynamic response that allows us to estimate and make inferences about the delay or latency irrespective of the amplitude. Our approach is very similar to the recent work of Henson *et al.* (2002). For a detailed comparison see the note added in proof (Appendix A.4).

We have chosen a model for the HRF in which the time axis of a fixed reference HRF is shifted by a single unknown parameter. We will show that our method is reasonably robust to misspecification of the form of the HRF. In other words, even if the time shift model is not correct (which remains to be investigated but is not the purpose of this paper), our method still gives a reasonably accurate estimate of the delay. Moreover our method is quite general and could be applied to any single parameter that modulates the HRF, such as the dispersion.

Finally, we have chosen a method of estimation which is fast (it takes only slightly longer than estimating the magnitude of the response), general (it can estimate different delays for different stimuli simultaneously), efficient (it is almost as good as the non-linear least-squares estimator), valid (it is roughly unbiased and has low root mean squared error) and reasonably robust to the assumed model for the HRF.

In Section 2 we shall develop the method, and in Section 3.1 we shall validate it on simulated data. In Section 3.2 we shall apply it to an fMRI experiment in pain perception. Some discussion and concluding remarks will be given in Sections 4 and 5. Finally, the complete `MATLAB` code for the method is available from `http://www.math.mcgill.ca/keith/fmristat`.

2 Methods

The first step in the statistical analysis of fMRI data is to build a model of how the data responds to an external stimulus. Suppose the external stimulus at time t is given by $s(t)$ and the fMRI response at a particular voxel is given by $x(t)$. For the pain experiment to be analyzed in Section 3.2, we might model the hot stimulus by 1 when the stimulus is on, and 0 when it is off (Figure 1(a)). The corresponding fMRI response is not instantaneous; there is a delay and blurring of the peak response by about 6 seconds. The simplest way of capturing this is to assume that the fMRI response depends on the external stimulus by convolution with an HRF $h(t)$ as follows:

$$x(t) = (s \star h)(t) = \int_0^\infty s(t-u)h(u)du,$$

where $\star$ is the convolution operator. Several models for $h(t)$ have been proposed. The simplest is a gamma function (Lange and Zeger, 1997) or a difference of two gamma functions to model the slight intensity dip after the response has fallen back to zero (Friston *et al.*, 1998). An example is the HRF available in SPM'96:

$$h(t) = (t/d)^a \exp(-(t-d)/b) - c(t/d')^{a'} \exp(-(t-d')/b'), \tag{1}$$

where t is time in seconds, $d = ab$ is the time to the peak, $d' = a'b'$ is the time to the under-shoot, and $a = 6$, $a' = 12$, $b = b' = 0.9$ seconds, and $c = 0.35$ (Glover, 1999) (Figure 1(b)). Its convolution $x(t)$ with the hot and warm stimuli $s(t)$ are shown in Figure 1(c). This is then subsampled at the n frame acquisition times $t_1, \ldots, t_n$ to give the response $x(t_i)$ at frame i. The linear model for the observation at frame i is:

$$Y_i = x(t_i)\beta + \varepsilon_i, \tag{2}$$

where the error ε_i has a Normal distribution with mean zero and standard deviation σ, $i = 1, \ldots, n$ (Friston *et al.*, 1995). Without loss of generality, we can assume that the errors in (2) are independent; correlated errors will be discussed in Section 2.4. Further responses and linear drift terms, such as a polynomial in time t_i or low-frequency cosine transform basis functions (SPM'99), can be added to the model (2), but for simplicity of presentation we shall just consider a single response and no drift terms. These can easily be added later (see Section 2.4).

2.1 Proposed time shifted model

So far we have assumed a fixed parametric form for the hemodynamic response function. Although the parameters are usually reasonably well known, it is still worth estimating these parameters. For some types of experiment, the parameters themselves, such as the delay, are of intrinsic interest for determining the chronology of brain activation.

The first step is to create a parametric family of HRF's. To do this, we start with a *reference* HRF $h_0(t)$ which should be chosen to be close to the true HRF, for example the gamma difference HRF (1). We then propose varying the delay by shifting the origin of the HRF by δ, to give

$$h(t; \delta) = h_0(t - \delta),$$

and the problem is to estimate δ. For practical purposes, we shall measure the delay d by the time to the first peak of the HRF, in other words $d = d_0 + \delta$, where d_0 is the time to the first peak of the reference HRF, for example $d_0 = 5.4$ seconds for the gamma difference HRF (1). An example of the time shifted model is shown in Figure 2(a).

2.2 The spectral estimator of the time shift

Friston *et al.* (1998) proposed capturing the shifted HRF $h(t; \delta)$ by linear combinations of just two basis functions $h_0(t)$ and

$$h_1(t) = \left.\frac{\partial h(t; \delta)}{\partial \delta}\right|_{\delta=0} = -\dot{h}_0(t),$$

where dot represents derivative with respect to time (see Figure 2(b)). If the time scale shift δ is small, that is the true HRF is close to the reference HRF, then we can approximate $h(t; \delta)$ by the first two terms of a Taylor series about $\delta = 0$:

$$h(t; \delta) \approx h_0(t) + h_1(t)\delta. \tag{3}$$

We can convolve the stimuli with $h_0(t)$ and $h_1(t)$, then estimate their coefficients by least squares. The ratio of these coefficients is the ratio estimator of δ. We might naturally ask if there is another choice of the two basis functions that can better approximate $h(t;\delta)$ over a range $-\Delta \leq \delta \leq \Delta$ (Figure 2(a)), in other words, basis functions $u_0(t), u_1(t)$ and coefficients $w_0(\delta), w_1(\delta)$ chosen so that

$$h(t;\delta) \approx u_0(t)w_0(\delta) + u_1(t)w_1(\delta). \tag{4}$$

In Appendix A.3 we show how to choose the basis functions in an optimal way using a singular value decomposition; the results are shown in Figure 2(b),(c). Letting

$$x_0 = s \star u_0, \quad x_1 = s \star u_1,$$

the linear model (2) then becomes

$$Y_i \approx x_0(t_i)w_0(\delta)\beta + x_1(t_i)w_1(\delta)\beta + \varepsilon_i, \tag{5}$$

which is now a linear model in $\gamma_0 = w_0(\delta)\beta$ and $\gamma_1 = w_1(\delta)\beta$. These can be estimated rapidly by least-squares as follows. Switching to matrix notation, let $\mathbf{Y} = (Y_1, \ldots, Y_n)'$, $\mathbf{x}_0 = (x_0(t_1) \ldots x_0(t_n))'$, $\mathbf{x}_1 = (x_1(t_1) \ldots x_1(t_n))'$ and $\mathbf{X} = (\mathbf{x}_0 \quad \mathbf{x}_1)$. The least squares estimators are

$$\begin{pmatrix} \widehat{\gamma}_0 \\ \widehat{\gamma}_1 \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \tag{6}$$

This gives estimates of two functions of the two unknown parameters β and δ. We can now equate the functions to their estimates:

$$w_0(\widehat{\delta})\widehat{\beta} = \widehat{\gamma}_0, \quad w_1(\widehat{\delta})\widehat{\beta} = \widehat{\gamma}_1, \tag{7}$$

and solve for $\widehat{\delta}$ and $\widehat{\beta}$. This can be done uniquely provided

$$r(\delta) = \frac{w_1(\delta)}{w_0(\delta)}$$

is a monotonic function of δ. If this is so our estimator of δ is the solution of

$$r(\widehat{\delta}) = \widehat{r} = \frac{\widehat{\gamma}_1}{\widehat{\gamma}_0}. \tag{8}$$

Clearly this estimator will be highly variable if $\widehat{\gamma}_0$ is small. Since $w_0(\delta) > 0$ (see Figure 2(c)), this will happen when β is small and the stimulus produces little response. In Appendix A.1 we show how to improve $\widehat{r}$ by finding its expectation, then correcting the bias by shrinking the estimator towards zero if little signal is present. To do this, we first define the T statistic for testing for nonzero magnitude, that is $\beta = 0$, or equivalently, $\gamma_0 = w_0(\delta)\beta = 0$. The residuals are

$$\mathbf{R} = \mathbf{Y} - \mathbf{X}\begin{pmatrix} \widehat{\gamma}_0 \\ \widehat{\gamma}_1 \end{pmatrix},$$

and the estimated variance is

$$\widehat{\sigma}^2 = \frac{\mathbf{R}'\mathbf{R}}{\nu}$$

where $\nu = n - 2$ is the residual degrees of freedom. The T statistic for the magnitude is

$$T_0 = \frac{\widehat{\gamma}_0}{\widehat{\mathrm{Sd}}(\widehat{\gamma}_0)} = \frac{\widehat{\gamma}_0}{\sqrt{V_0}\widehat{\sigma}}, \tag{9}$$

where V_0 is the first diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$. The proposed corrected estimator of r is

$$\widehat{r}_{\mathrm{C}} = \frac{\widehat{r}}{1 + \frac{1}{T_0^2}}, \tag{10}$$

and our proposed (corrected) spectral estimator of δ is

$$\widehat{\delta} = r^{-1}(\widehat{r}_{\mathrm{C}}). \tag{11}$$

We can see that when there is little evidence for a stimulus response and so T_0 is really small, $\widehat{r}_{\mathrm{C}}$ is shrunk to zero. Since $r(0) = 0$ (see Figure 2(c)), then $\widehat{\delta}$ is also shrunk to zero where there is little evidence for activation.

The spectral method can be applied to any single parameter that modifies the HRF, such as a shift of scale rather than a shift of origin. It can also be extended to estimate two or more parameters simply by adding more basis functions, one for each extra parameter. Provided the mapping from coefficients to parameters is one-to-one, the coefficients can be inverted to estimate the parameters, though in practice this inversion can be time consuming. Is there anything to be gained by adding yet more basis functions, in order to better approximate the HRF? The answer is that of course the approximation will be better, but now one will now have more equations than unknowns, forcing us to resort to iterative non-linear methods to estimate the parameters, something that our method is striving to avoid.

2.3 Inference about the delay

So far we have only looked at δ, the shift parameter. The parameter of most interest is the delay d to the first peak. Let d_0 be the delay for the reference HRF h_0. Then

$$d = d_0 + \delta.$$

Our estimator of d is

$$\widehat{d} = d_0 + \widehat{\delta}$$

and its estimated standard deviation $\widehat{\mathrm{Sd}}(\widehat{d})$ is given in Appendix A.2, equation (13). We can now test that d takes a specified value d^* in the usual way using a T statistic

$$T^* = \frac{\widehat{d} - d^*}{\widehat{\mathrm{Sd}}(\widehat{d})}.$$

The degrees of freedom ν is the residual degrees of freedom of the linear model that includes the pair of terms x_0 and x_1 for each stimulus, together with the drift terms. If $d = d^*$, then T^* has an approximate t distribution with ν degrees of freedom.

However for the special case of testing that d is equal to the reference delay d_0 there is an exact test. Note that if $d = d_0$ then $\delta = 0$ and so $\gamma_1 = w_1(\delta)\beta = 0$ (see Figure 2(b)). Hence the test can be based on the T statistic

$$T_1 = \frac{\widehat{\gamma}_1}{\widehat{\mathrm{Sd}}(\widehat{\gamma}_1)}$$

where $\widehat{\mathrm{Sd}}(\widehat{\gamma}_1) = \sqrt{V_1}\widehat{\sigma}$ and V_1 is the second diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$. If $d = d_0$ then T_1 has an exact t distribution with ν degrees of freedom.

A more interesting question is to compare two delays. Comparing delays at two different well separated voxels is straightforward. Because we can assume that distant voxels are independent, then to test that two delays d_1 and d_2 are equal, we can use the test statistic

$$\frac{\widehat{d}_1 - \widehat{d}_2}{\sqrt{\widehat{\mathrm{Sd}}(\widehat{d}_1)^2 + \widehat{\mathrm{Sd}}(\widehat{d}_2)^2}}$$

which will have an approximate t distribution with ν degrees of freedom if the two delays are equal. The most interesting question is to compare delays at the *same* voxel, or in general, an arbitrary linear contrast in the delays specified by $\mathbf{cd}$, where $\mathbf{c}$ is a row vector of the contrasts and $\mathbf{d}$ is a column vector of the delays. To do this we need the covariances between the delays, as well as their variances, which are given in Appendix A.2. The estimated standard deviation of the contrast is

$$\widehat{\mathrm{Sd}}(\mathbf{c}\widehat{\mathbf{d}}) \approx \sqrt{\mathbf{c}\widehat{\mathrm{Var}}(\widehat{\mathbf{d}})\mathbf{c}'}.$$

The test statistic for the contrast is

$$\frac{\mathbf{c}\widehat{\mathbf{d}}}{\widehat{\mathrm{Sd}}(\mathbf{c}\widehat{\mathbf{d}})}$$

which will once again have an approximate t distribution with ν degrees of freedom if $\mathbf{cd} = 0$.

2.4 Adding more stimuli and drift terms

So far we have only considered the case of a single stimulus and no drift terms. More stimuli with or without different shift parameters δ can be added, and drift terms can be incorporated as extra columns of $\mathbf{X}$. We are thus able to produce different estimators for the delay for each stimulus, and remove drift at the same time. For correlated errors with known covariance matrix $\mathbf{V}\sigma^2$, this method can be used by first transforming the model (5) by multiplying through both sides by $\mathbf{V}^{-\frac{1}{2}}$. This then produces independent errors ('white noise') as required. Autocorrelation parameters can be estimated from the residuals to estimate $\mathbf{V}$, which can then be used to whiten the data and the model by multiplying both by $\mathbf{V}^{-\frac{1}{2}}$ before re-estimating all the parameters. For a method of fitting such a model, see Worsley *et al.* (2002).

2.5 Estimating the magnitude β

Solving (7) for β, a natural estimator of β is $\widehat{\gamma_0}/w_0(\widehat{\delta})$. Simulations show that this estimator of the magnitude is reasonably accurate for a wide range of delays, but it is very sensitive to the dispersion of the HRF. A better estimator of the magnitude should also attempt to estimate the dispersion as well as the shift. This will be the subject of future research. As a compromise, we propose to estimate the magnitude by first scaling $u_0(t)$ by dividing by $\int u_0(t)dt$ (in fact both $w_0(\delta)$ and $w_1(\delta)$ in Figure 2(c) have been multiplied by $\int u_0(t)dt$ to compensate for this). Since now $u_0(t)$ integrates to 1, then for block designs with long blocks the coefficient of $x_0(t)$ is the signal magnitude. We then estimate the signal magnitude β by $\widehat{\beta} = \widehat{\gamma}_0$. As already pointed out, this estimator will be unbiased for block designs with long blocks. However for event related designs, where knowledge of the shape of the HRF is crucial, this estimator is slightly biased. From Figure 2(c) the resulting $w_0(\delta)$ is slightly greater than 1 if the shift is less than 2 seconds, suggesting that for event related designs the magnitude will be slightly over estimated in this region, and slightly under estimated if the shift is more than 2 seconds. Altering the dispersion has a greater effect: a change of ± 2 seconds in the dispersion of the HRF produces about a 20% change in the estimated magnitude. Note that neither a shift in the delay nor a change in the dispersion has any effect on the bias of $\widehat{\beta}$ for block designs with long blocks.

3 Results

3.1 Simulations

We validated the corrected estimators by simulation of noise distributions and stimuli typically found in fMRI data sets. We looked at bias, root mean squared error, standard deviation and estimated standard deviation to see if the model parameters derived analytically comply with the experimentally obtained results.

To assess the parameter estimators, we used simulated fMRI data sets with 120 frames, separated by 3 seconds, excluding the first two frames. The block design of 3 frames rest, 3 frames hot stimulus, 3 frames rest, 3 frames warm stimulus, was repeated 10 times so we have 120 frames in total (see Figure 1(a)). The reference HRF was the difference of two gamma functions (1). We added polynomial covariates of degree 3 in the frame times to the design matrix to remove the drift. Note that these drift terms are not convolved with the HRF. We simulated the fMRI experiment 2000 times with AR(1) errors and an autocorrelation of 0.3, typical of the values in real data. We only chose the hot stimulus for delay estimation.

We investigated the properties of three estimators:

1. The ratio estimator using $h_0(t)$ and $h_1(t)$ as HRF basis functions, and estimating the shift by the ratio of their coefficients.
2. The corrected ratio estimator obtained by dividing the ratio estimator by $1 + 1/T_0^2$, where T_0 is the T statistic for the magnitude of the stimulus (9).
3. The spectral estimator $\widehat{\delta}$ obtained by using optimally chosen basis functions and the relationship between the (corrected) ratio of their coefficients and the true shift.

We varied the true shift from -4.5 to 4.5 seconds, the standardised magnitude

$$\tau = \frac{\beta}{\mathrm{Sd}(\widehat{\beta})} = \frac{\beta}{\sqrt{V_0}\sigma},$$

(approximated by the T statistic T_0), from 1 to 10 (Figure 3) and the dispersion of the HRF from 2.6 to 10.4 seconds (by multiplying the time scale of the HRF by factors from 0.5 to 2 about the first peak with a width of 5.2 seconds) (Figure 4). The dispersion was measured by $\sqrt{8\log 2}b\sqrt{a}$, which is the FWHM of a Gaussian density with the same peak curvature as the first gamma density.

We expect the methods for estimating the delay to break down when the signal to noise ratio is too small to detect the signal in the first place. This is confirmed by Figure 3(a), which shows that the simple ratio estimator is highly unstable if the magnitude of the signal is small, resulting in a large RMSE (Figure 3(d)). Correcting the ratio estimator by shrinking it to zero when the magnitude is small overcomes these problems (Figures 3(b,e)) but there is still large bias for shifts greater than 3 seconds. The spectral estimator controls the bias to within 0.5 seconds even for shifts as large as 4.5 seconds, and it has lower RMSE (Figures 3(c,f)). The estimated standard deviation of the spectral estimator is within 5% of the true standard deviation when the magnitude is large (Figures 3(g-i)); the overestimated standard deviations for small magnitudes result in more conservative inference about the shift, which is not too serious.

So far in Figure 3 we have varied the delay of the HRF, keeping the shape of the HRF equal to that of the reference HRF. In Figure 4 we try varying the dispersion of the HRF about its peak to assess robustness of our estimators, keeping the standardised magnitude of the signal at $\tau = 6$ (marked by a dotted horizontal line on Figure 3). We see that all methods of estimating the shift are remarkably robust to the dispersion, although there is still some bias for large shifts and extreme dispersions (Figures 3(a-c)). The RMSE is smallest when the dispersion matches that of the reference HRF (5.2 seconds), indicated by a horizontal dotted line (Figures 3(d-f)). Overall the spectral estimator appears to perform slightly better than the other two. Its estimated standard deviation is within 5% of the truth over nearly all the shifts and dispersions.

Overall we conclude that the spectral estimator is reasonably accurate at detecting shifts up to ± 4.5 seconds if the standardised magnitude is at least 4, even if the true dispersion varies by a factor of 2 from that assumed.

3.2 Illustrative example

Using the same experimental protocol as in Section 3.1, data were collected from several scans on a group of subjects, but we shall just analyse the data from a single scan on a single subject (Chen *et al,* 2000). As we mentioned before, the reference HRF was modeled as the difference of two gamma density functions with a delay of 5.4 seconds to the peak (1). First, the T statistic T_0 for the magnitude is shown in Figure 5(a). It is clear that there is a significant area in the middle region where the T statistic is above 4.5 (red-white color), indicating that there is strong signal in this region (the center of the 3×3 array in Figure 6 shows a magnification of this region). Figure 5(b) shows the T statistic T_1 for the time shift.

Note that it is close to zero in the red-white region, indicating no evidence of a difference between the delay and the reference delay. Figure 5(c) shows the estimated delay $\widehat{d}$, which is close to the reference delay $d_0 = 5.4$ seconds, as expected from the test in Figure 5(b). Furthermore the standard deviation of the estimated delay (Figure 5(d)) was $\widehat{\mathrm{Sd}}(\widehat{d}) \approx \pm 0.6$ seconds over almost all the region.

To check the success of the method, we used different reference HRF's: we shifted the original gamma difference HRF by ± 3 seconds to give a reference delay of $5.4 + \pm 3$ seconds, and we varied the dispersion by a factor of two about the first peak. If our method is successful, then we should still estimate the delay as about 5.4 seconds even using these different choices of reference HRF. The results are shown in Figure 6 where we only concentrate on the region of activation in Figure 5. In (a) we see that the T statistic for the magnitude is highest when the peak of the reference HRF is close to the anticipated HRF, in other words, the stimulus is easiest to detect when the reference HRF is close to the true one. In (c) we note that no matter what reference HRF is chosen, the estimated delay remains at about 5.4 seconds. This is confirmed in (b) where we see that the shift is not different from zero when the reference delay is 5.4, but it is greater than zero when the reference delay is less than 5.4 seconds, and less than zero when the reference delay is greater than 5.4 seconds. In (d) we note that the standard error of the estimated delay is low when the reference HRF is close to the true one. This suggests that our method for estimating the delay is reasonably robust to the choice of reference HRF.

Figures 7 and 8 show the delay superimposed on the region of significant stimulus. The range of delays is from about 4.5 to 6.5 seconds, with a standard deviation of about 0.6 seconds. These are not significantly different from the reference delay of 5.4 seconds. Overall our conclusion is that there is little evidence in this data for a delay that is different from 5.4 seconds. There was also no evidence for a difference in delay between the hot and warm stimuli.

4 Discussion

4.1 On the difficulties of the interpretation of the GLM

First, we would like to stress that this work casts some light on some of the often not well enough acknowledged or documented difficulties of the analysis of functional data with the general linear model. Indeed, although linear models are both simple and flexible, and their resolutions comply with the speed requirements that are found in the analysis of large time series, the construction of the model is generally difficult. The choice of how many and which regressors to include in the model is often ad hoc (the functional imaging community lacks verification and model selection tools), and the results may well be misinterpreted. A first source of possible misinterpretation due to correlations of the regressors constituting the GLM have already been pointed out in Andrade *et al.* (1999).

There is another possible source of confusion even in the case of no correlation in the model. Typically, the comparison of the magnitude of two responses, when the model includes derivatives in time to accommodate for differential delays of those responses, is biased if indeed there is a difference in the delay of the responses. This has probably been very

often overlooked when analysing data. The magnitude of the bias of course depends on the difference of the delays, but it can be shown that for a delay around one second, the bias is around 7-8% of the magnitude, a figure comparable with the magnitude of changes observed in fMRI datasets. Of course, not modeling the delay does not solve the problem but only makes it impossible to look at and to solve. This will be the subject of future research.

To eliminate ambiguities of the linear model, one could resort to non linear fits of models that explicitly parametrise the delay, amplitude or other parameters. First, as mentioned before, these procedure are computationally costly, and the computation cost might well be itself a limit when analysing a group of subjects (often 10 to 20) with often more than a thousand time points per subject. Second, the non linear fitting in the presence of noise does not always converge to the global minimum and the computation is therefore often limited to the voxels that show an effect. The selection of those voxels can be done during the non linear fit but this may require an arbitrary threshold on the energy to be minimised. In the method proposed here, we solve for this by including in the corrected delay estimator the statistics that reflects the evidence for a response (see (10)).

4.2 On the shifted HRF model

The computations presented in this paper are based on a model of the HRF that shifts with time. A drawback of this model is that if the shift is negative then responses in the future can influence BOLD signal in the present. However, we usually anticipate positive shifts (longer delays) so this may not be a problem. We do not propose experimental data and a formal test to validate this model (although some data seem to show such an effect) since this would go beyond the purpose of this work. However, it has to be emphasised that the method proposed is reasonably robust to other forms of the HRF, including change in dispersion. It also extends to any single parameter model – we only need to replace u_0 and u_1 by the SVD of the spectrum of HRF's with respect to that parameter. If the shifted HRF model is shown to be less appropriate than another model – something that still has to be demonstrated – the basis for accurate estimators of both the delay and the magnitude would be found in previous sections. Interestingly, one of the earliest models proposed for the HRF in Friston et al. (1995) was of the form of a Poisson distribution in which the dispersion varies as the square root of the delay.

While a non-parametric estimation of the HRF is feasible, this estimation is noisy and is likely to generally require temporal regularisation. These operations are costly to perform at the voxel level, especially if several types of stimuli are considered for estimation. Extraction of parameters from a non-parametric form could be performed after a non-parametric estimation of the HRF but a parametric model would still be required to extract the delay.

4.3 On the ultimate goal: to retrieve the neuronal signal

Extracting various parameters from the HRF does inform on the neural activity only under some strong assumptions. The comparison of the delay or the magnitude of the response is most likely to make sense only at the same position in the brain or else it may only reflect differences in the microvasculature system across cerebral regions. Today, the amount of information on neural activity that can be extracted from the BOLD response is not

known and this question (together with the method used to extract this information) is an active field of research. Typically, if the HRF is known exactly or with great precision, not changing with time (stationarity), and the BOLD response is formed by convolution of the neural response with this function (linearity), one could deconvolve the measured BOLD signal to find the neural activity. However, non-linearities have been demonstrated by several authors (Boynton *et al.*, 1996; Glover *et al.*, 1999; Vazquez & Noll, 1998; Birn *et al.*, 2001) and current research investigates how much of these are due to non-linearities of the transfer from the stimulus to the neural response and how much are due to the transfer from the neural to the hemodynamic signal (see for instance Buxton *et al.* (2001)). Non-stationarities with time have yet to be investigated.

It is likely that questions related to the nature of the coupling between stimulus/neural and response/vascular response will find their answers with techniques that use concurrent recordings of electrical and metabolic responses, such as implanted electrodes in epileptic patients or EEG recording in the MRI scanner. At the moment, given our knowledge of both the shape and the characteristics with respect to the neural inputs of the HRF, little can be inferred on the neural activity from the BOLD signal with certitude and many different neural inputs can lead to almost identical BOLD signals. Yet, a precise and rapid estimation of the delay of the HRF is important since for well controlled experimental designs for which the neural response is thought to be modulated in terms of delay and magnitude, the HRF parameters are likely to reflect actual neural changes, as confirmed by a recent study using optical imaging (see Gratton *et al.*, 2001). In fact, to investigate if the variation of those parameters actually represents neural changes, accurate estimation is needed.

5 Conclusions

We have implemented an efficient, robust algorithm for estimating the delay of the response in fMRI data. The `MATLAB` program (`http://www.math.mcgill.ca/keith/fmristat`) takes about 4 minutes to analyse one fMRI data set. We then used this algorithm for statistical inference of the delay of the response to a painful heat stimulus. From the data set analyzed, we found significant positive responses in three main areas (Figures 7 and 8). In these areas, the estimated delay was around 5.4 seconds and it varied little with the reference HRF that was used. This last point confirms the success of the method.

A Appendix

A.1 Details of the corrected estimator

It can be shown that the expected value of the ratio estimator, the ratio of two Gaussian random variables, does not exist, since the ratio becomes infinite when the denominator approaches zero. Nevertheless, we shall approximate the ratio locally by a Taylor series, then take expectations. Equating this to the true ratio will lead to a new estimator whose expectation does exist, and is close to the true value. Assuming $\widehat{\gamma}_0$ and $\widehat{\gamma}_1$ are roughly independent (true for an event related design with no autocorrelation and no drift terms,

since $\mathbf{u_0}$ and $\mathbf{u_1}$ are orthogonal) then

$$\mathrm{E}(\widehat{r}) \approx \mathrm{E}(\widehat{\gamma}_1)\mathrm{E}(1/\widehat{\gamma}_0).$$

Since $\widehat{\gamma}_0$ and $\widehat{\gamma}_1$ are least squares estimators they are unbiased so that $\mathrm{E}(\widehat{\gamma}_0) = \gamma_0$ and $\mathrm{E}(\widehat{\gamma}_1) = \gamma_1$. Now let $f(u) = 1/u$, then expanding this as a Taylor series about γ_0 and taking expectations, we have:

$$\mathrm{E}(1/\widehat{\gamma}_0) = \mathrm{E}(f(\widehat{\gamma}_0)) \approx f(\gamma_0) + \mathrm{Var}(\widehat{\gamma}_0)\ddot{f}(\gamma_0)/2.$$

Combining the above results we get the following formula to estimate r:

$$\begin{aligned} \mathrm{E}(\widehat{r}) &\approx \frac{\gamma_1}{\gamma_0}\left(1 + \frac{\mathrm{Var}(\widehat{\gamma}_0)}{\gamma_0^2}\right) \\ &\approx r\left(1 + \frac{1}{\tau^2}\right), \end{aligned}$$

neglecting terms in r^3 and higher. We can estimate τ by T_0 from (9). This suggests that a better estimator of r is:

$$\widehat{r}_{\mathrm{C}} = \widehat{r}\left(1 + \frac{1}{T_0^2}\right)^{-1},$$

which now approaches 0 as $\widehat{\gamma}_0$ approaches 0, so its expectation exists, and is within 2% of the true ratio if $|\tau| > 4$.

A.2 Estimating the variance and covariance

Let

$$g(v, u) = r^{-1}\left(\frac{u/v}{1 + (s/v)^2}\right),$$

where $s = \widehat{\mathrm{Sd}}(\widehat{\gamma}_0)$, so that $\widehat{\delta} = g(\widehat{\gamma}_0, \widehat{\gamma}_1)$. Then by standard linear Taylor series approximations, ignoring the variability in s, the estimated variance of $\widehat{\delta}$ is

$$\widehat{\mathrm{Var}}(\widehat{\delta}) \approx \dot{\mathbf{g}}(\widehat{\gamma}_0, \widehat{\gamma}_1)\widehat{\mathrm{Var}}\begin{pmatrix} \widehat{\gamma}_0 \\ \widehat{\gamma}_1 \end{pmatrix}\dot{\mathbf{g}}(\widehat{\gamma}_0, \widehat{\gamma}_1)', \tag{12}$$

where

$$\widehat{\mathrm{Var}}\begin{pmatrix} \widehat{\gamma}_0 \\ \widehat{\gamma}_1 \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\hat{\sigma}^2.$$

Then

$$\dot{\mathbf{g}}(v, u) = \left(\frac{u(s^2 - v^2)}{(v^2 + s^2)^2}, \quad \frac{v}{v^2 + s^2}\right)/\dot{r}(g(v, u)),$$

and so, letting $a_0 = 1 + 1/T_0^2$,

$$\dot{\mathbf{g}}(\widehat{\gamma}_0, \widehat{\gamma}_1) = \left(\frac{\widehat{r}\,(a_0 - 2)}{\widehat{\gamma}_0 a_0^2}, \quad \frac{1}{\widehat{\gamma}_0 a_0}\right)/\dot{r}(\widehat{\delta}).$$

The estimated variance of the delay $\widehat{d} = d_0 + \widehat{\delta}$ is

$$\widehat{\mathrm{Sd}}(\widehat{d})^2 = \widehat{\mathrm{Var}}(\widehat{d}) \approx \widehat{\mathrm{Var}}(\widehat{\delta}). \tag{13}$$

The estimated covariance between two delays $\widehat{d}_1$ and $\widehat{d}_2$ at the same voxel is

$$\widehat{\mathrm{Cov}}(\widehat{d}_1, \widehat{d}_2) \approx \dot{\mathbf{g}}(\widehat{\gamma}_{01}, \widehat{\gamma}_{11}) \begin{pmatrix} \mathrm{Cov}(\widehat{\gamma}_{01}, \widehat{\gamma}_{02}) & \mathrm{Cov}(\widehat{\gamma}_{01}, \widehat{\gamma}_{12}) \\ \mathrm{Cov}(\widehat{\gamma}_{11}, \widehat{\gamma}_{02}) & \mathrm{Cov}(\widehat{\gamma}_{11}, \widehat{\gamma}_{12}) \end{pmatrix} \dot{\mathbf{g}}(\widehat{\gamma}_{02}, \widehat{\gamma}_{12})'.$$

The elements of the inner 2×2 matrix come from the corresponding elements of $(\mathbf{X}'\mathbf{X})^{-1}\widehat{\sigma}^2$, where $\mathbf{X}$ now includes the pair of columns $\mathbf{x}_0$ and $\mathbf{x}_1$ for each stimulus, together with columns for the drift.

A.3 Details of the spectral estimator

How do we choose the basis functions? One method is to choose $h(t;\delta)$ for two fixed values of δ, say $\pm\delta*$ where $\delta* = \Delta/2$. This has the advantage that the approximation (4) is exact for these two values of δ, and perhaps fairly accurate for the others. Equivalently, we could use the average and the difference of these two functions, equivalent to the average and its numerical derivative. As the value of $\delta* \to 0$, this is equivalent to using h_0 and h_1 as the basis functions, leading to the ratio estimator. Thus the ratio estimator is likely to be very accurate in the neighbourhood of $\delta = 0$.

A better criterion might be to minimise the proportion of variability of $h(t;\delta)$ accounted for by its approximation (4), integrated over all values of δ, rather than at two fixed values of δ as just described. To do this in practice, form a matrix $\mathbf{H}$ by sampling $h(t;\delta)$ at equally spaced values of t (columns) and δ (rows) in the range $-\Delta \le \delta \le \Delta$. Then find the singular value decomposition (svd):

$$\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}' \approx \mathbf{u}_0 s_0 \mathbf{v}_0' + \mathbf{u}_1 s_1 \mathbf{v}_1'$$

where $\mathbf{U} = (\mathbf{u}_0\ \mathbf{u}_1\ \cdots)$ and $\mathbf{V} = (\mathbf{v}_0\ \mathbf{v}_1\ \cdots)$ are orthonormal matrices and $\mathbf{S} = \mathrm{diag}(s_0 \ge s_1 \ge \cdots)$. The success of the approximation is measured by the proportion of variability explained, equal to $(s_0^2+s_1^2)/(s_0^2+s_1^2+\cdots) = 0.87$ for the gamma difference HRF in the range $-4.5 \le \delta \le 4.5$ (whereas only 0.75 is explained by h_0 and h_1). The functions $w_0(\delta), w_1(\delta)$ are then approximated by the elements of $s_0\mathbf{v}_0, s_1\mathbf{v}_1$. Note that if we had chosen just two columns for $\mathbf{H}$ then this method would be identical to that in the previous paragraph.

We notice straight away a strong similarity between $\mathbf{u}_0$, $\mathbf{u}_1$ and the normalised functions $h_0(t)$, $h_1(t)$ (Figure 2(b)). Note also the symmetry of $\mathbf{v}_0$ and $\mathbf{v}_1$ as functions of δ (Figure 2(c)). There is a simple reason for this: $\mathbf{H}'\mathbf{H}$ is a Toeplitz matrix, whose eigenvectors $\mathbf{V}$ are either symmetric or antisymmetric about their middle element. A row or column of the Toeplitz matrix is just the autocorrelation function of h_0, whose eigenfunctions are the Fourier basis functions. If the spectrum is unimodal, then the eigenfunctions corresponding to the two largest eigenvalues are the lowest frequency sine and cosine waves. This is precisely what we see in Figure 2(c), altered by the fact that the range of δ is truncated to $[-\Delta, \Delta]$. The first eigenvector $\mathbf{v}_0$ resembles a cosine wave, the second $\mathbf{v}_1$ a sine wave. These are close to a constant and linear function of δ as given by the Taylor series expansion, which explains why the corresponding $\mathbf{u}_0$ and $\mathbf{u}_1$ resemble $h_0(t)$ and $h_1(t)$.

The key to the success of this method is the monotonicity of $r(\delta)$. Fortunately Figure 2(b) confirms that this is the case, and moreover it is quite close to a straight line, the principle behind the ratio estimator. Monotonicity of $r(\delta)$ breaks down if the range of δ is increased beyond ± 5.6 seconds, making the method inoperative. Since as noted above this function is the ratio of two functions that resemble a sine and cosine, it was approximated by a scaled tangent function (dashed line on Figure 2(c), fitted by non-linear least squares) which allows us to invert it rapidly using the arctangent function.

A.4 Note added in proof: Comparison with the method of Henson et al. (2002)

Since the final revision of this paper, a very similar method has been published by Henson *et al.* (2002). This paper also proposes a simple method of estimating the delay using two basis functions and inverting the relationship between the ratio of the coefficients and the true delay. There are three main differences. First, Henson *et al.* (2002) use Taylor series basis functions (3) rather than our optimally chosen basis functions (4). Second, in Henson *et al.* (2002) the ratio of the coefficients is inverted directly as in (8) without our shrinkage as in (10). Henson *et al.* (2002) use a logistic approximation here, whereas we prefer an arctangent approximation, but the difference is negligible. Simulations show that there is little difference between the two methods except that our shrinkage estimator yields better results when the signal magnitude is small. Thirdly, Henson *et al.* (2002) rely on replications of the experiment to estimate the standard error, whereas we can obtain a standard error directly from a single replication, as in (13).

References

Andrade A., Paradis A.-L., Rouquette S., & Poline J.-B. (1999). Ambiguous results in functional neuroimaging data analysis due to covariate correlation. *NeuroImage*, **10**:484-486.

Birn, R.M., Saad, Z.S. & Bandettini, P.A. (2001). Spatial heterogeneity of the nonlinear dynamics in the fMRI BOLD response. *NeuroImage*, **14**:817-826.

Boynton, G.M., Engel, S.A., Glover, G.H. & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging data in human V1. *Journal of Neuroscience*, **16**:4207-4221.

Buxton, B. R., Liu T.T., & Wong E. C. (2001). Nonlinearity of the hemodynamic response: Modeling the neural and BOLD contributions. *Proc. Intl. Soc. Reson. Med.*, 1164.

Chen, J-I., Ha, B., Bushnell, M.C., Pike, B. & Duncan, G.H. (2000). Differentiating pain- and innocuous mechanical-related activation of human S1 cortex using temporal analysis of fMRI. *In preparation.*

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J-B., Frith, C.D. & Frackowiak, R.S.J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, **2**:189-210.

Friston, K.J., Fletcher, P., Josephs, O., Holmes, A.P., Rugg, M.D., & Turner, R. (1998). Event-related fMRI: Characterising differential responses. *NeuroImage*, **7**:30-40.

Glover, G.H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, **9**:416-429.

Gratton G, Goodman-Wood M. R., & Fabiani M (2001). Comparison of neuronal and hemodynamic measures of the brain response to visual stimulation: an optical imaging study. *Human Brain Mapping*, **13**:13-25.

Henson, R.N.A., Price, C.J., Rugg, M.D., Turner, R. & Friston, K.J. (2002). Detecting latency differences in event-related BOLD responses: application to words versus non-words and initial versus repeated face presentations. *NeuroImage*, **15**:83-97.

Lange, N. & Zeger, S.L. (1997). Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging (with Discussion). *Applied Statistics*, **46**:1-29.

Kruggel, F. & von Cramon, D.Y. (1999). Temporal properties of the hemodynamic response in functional MRI. *Human Brain Mapping*, **8**:259-271.

Kruggel, F., Zysset, S. & von Cramon, D.Y. (2000). Nonlinear regression of functional MRI data: an item recognition task study. *NeuroImage*, **12**:173-183.

Purdon, P.L., Solo, V., Weisskoff, R.M. & Brown, E. (2001). Locally regularized spatiotemporal modeling and model comparison for functional MRI. *NeuroImage*, **714**:912-923.

Rajapakse, J.C., Kruggel, F., Maisog, J.M. & von Cramon, D.Y. (1998). Modeling hemodynamic response for analysis of functional MRI time-series. *Human Brain Mapping*, **6**:283-300.

Saad, Z.S., Ropella, K.M., Cox, R.W. & DeYoe, E.A. (2001). Analysis and use of fMRI response delays. *Human Brain Mapping*, **13**:74-93.

Seber, G.A.F. & Wild, C.J. (1989) *Nonlinear Regression.* New York: Wiley.

Thierry, G., Boulanouar, K., Kherif, F., Ranjeva, J-P. & Démonet, J-F. (1999) Temporal sorting of neural components underlying phonological processing. *NeuroReport*, **10**: 2599-2603.

Vazquez, A.L. & Noll, D.C. (1998). Nonlinear aspects of the BOLD response in functional MRI. *NeuroImage*, **7**:108-18.

Worsley, K.J., Liao, C., Aston, J., Petre, V., Duncan, G.H., Morales, F. & Evans, A.C. (2002). A general statistical analysis for fMRI data. *NeuroImage*, **15**:1-15.

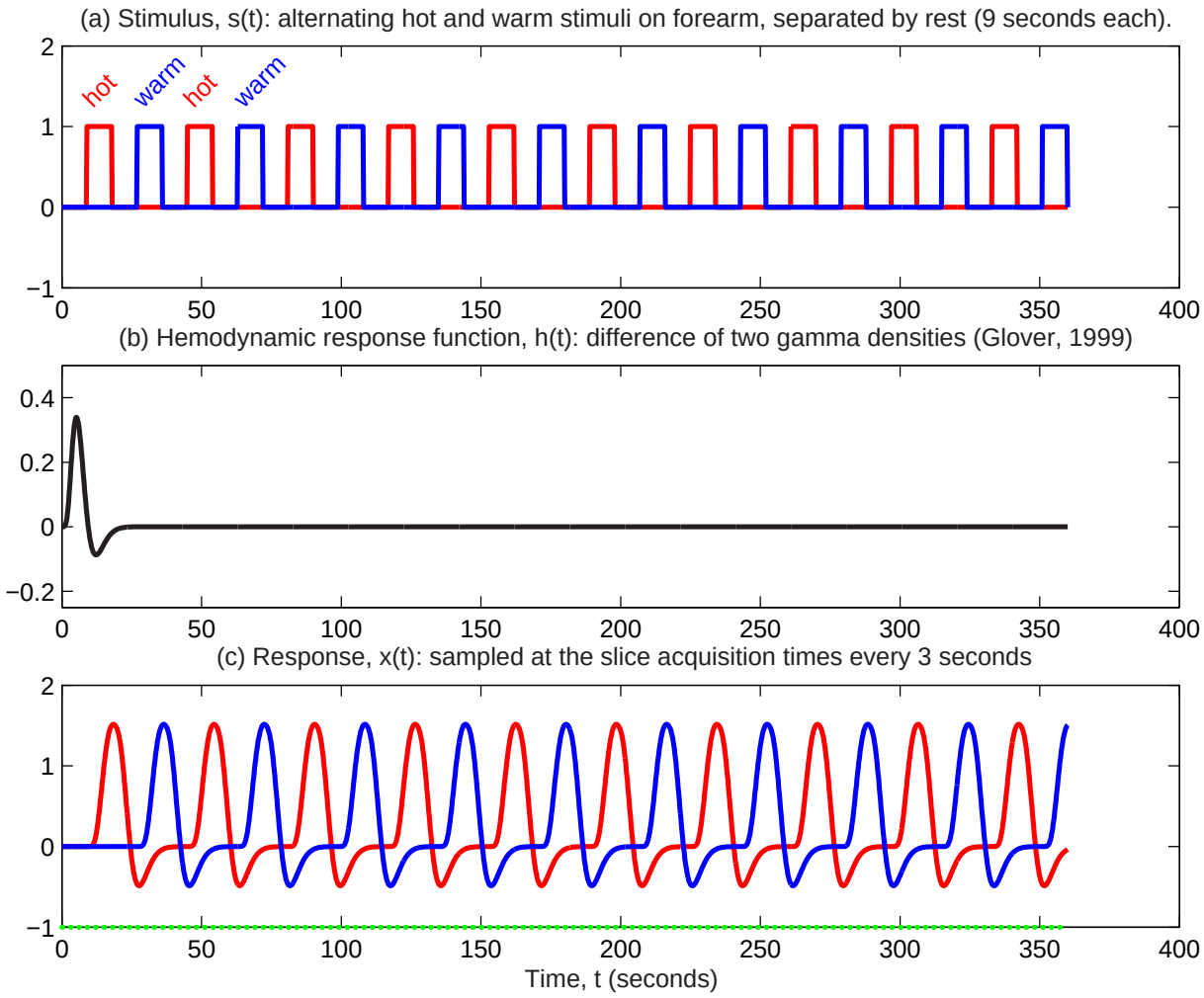

Figure 1: (a) The hot and warm stimuli $s(t)$, (b) the hemodynamic response function $h(t)$ and (c) its convolution with $s(t)$ to give the response $x(t)$. The time between frames is TR=3s, so $x(t)$ is then subsampled at the $n = 118$ frame acquisition times $t_i = 3i$ to give the response $x(t_i)$ at time index $i = 1, \ldots, n$.

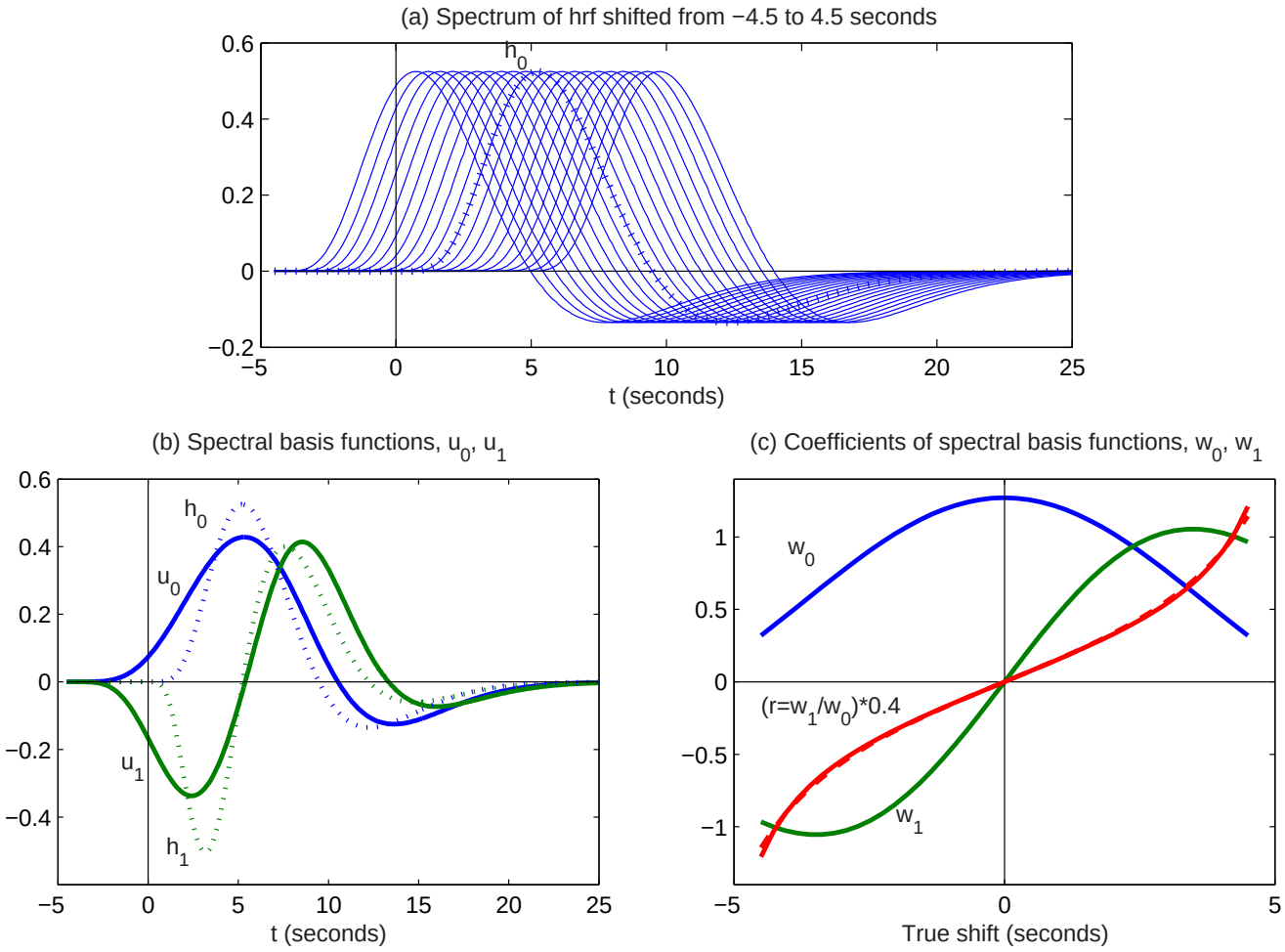

Figure 2: (a) HRFs shifted by $-4.5 \leq \delta \leq 4.5$ seconds; (b) the spectral basis functions (u_0, u_1) found by singular value decomposition (svd) of the HRFs in (a), compared to the Taylor series basis functions (h_0, h_1); (c) the corresponding coefficients of the spectral basis functions (w_0, w_1) as functions of the shift parameter, δ. The ratio $r = w_1/w_0$ (multiplied by 0.4) is monotonic in δ, so we can invert it to find the spectral estimator $\widehat{\delta}$. This is particularly rapid if we use a scaled tangent function approximation (dashed line).

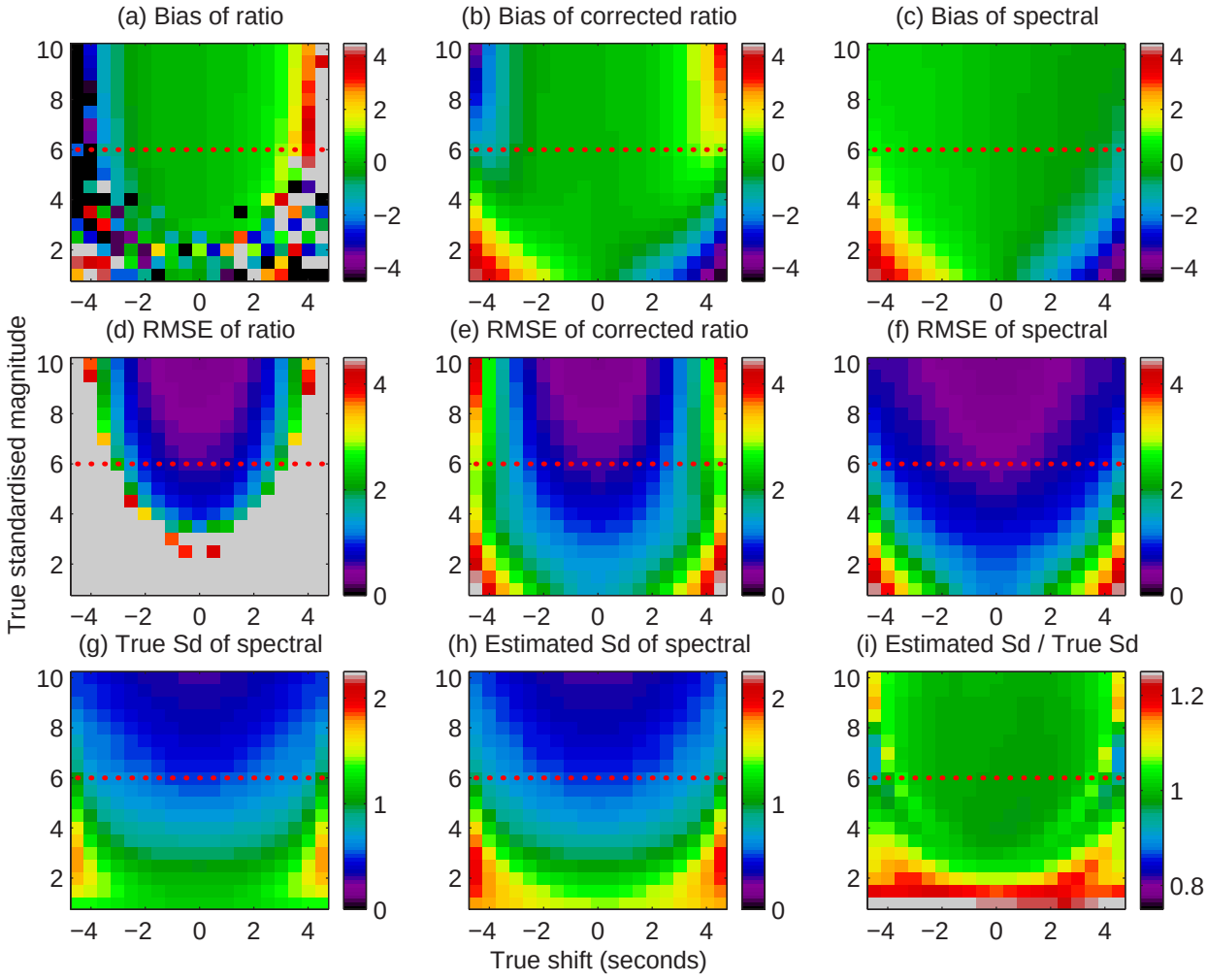

Figure 3: Bias and root mean squared error (RMSE) of the three shift estimators: the ratio estimator, the corrected ratio estimator, and the spectral estimator $\widehat{\delta}$, based on 2000 simulations of the hot stimulus, as a function of the true shift δ and the true standardised magnitude τ: (a-c) the bias; (d-f) the root mean squared error. The spectral estimator appears to be the best. (g) The true standard deviation of the spectral estimator, (h) its estimated standard deviation using the method in Section A.2, and (i) the ratio of the two.

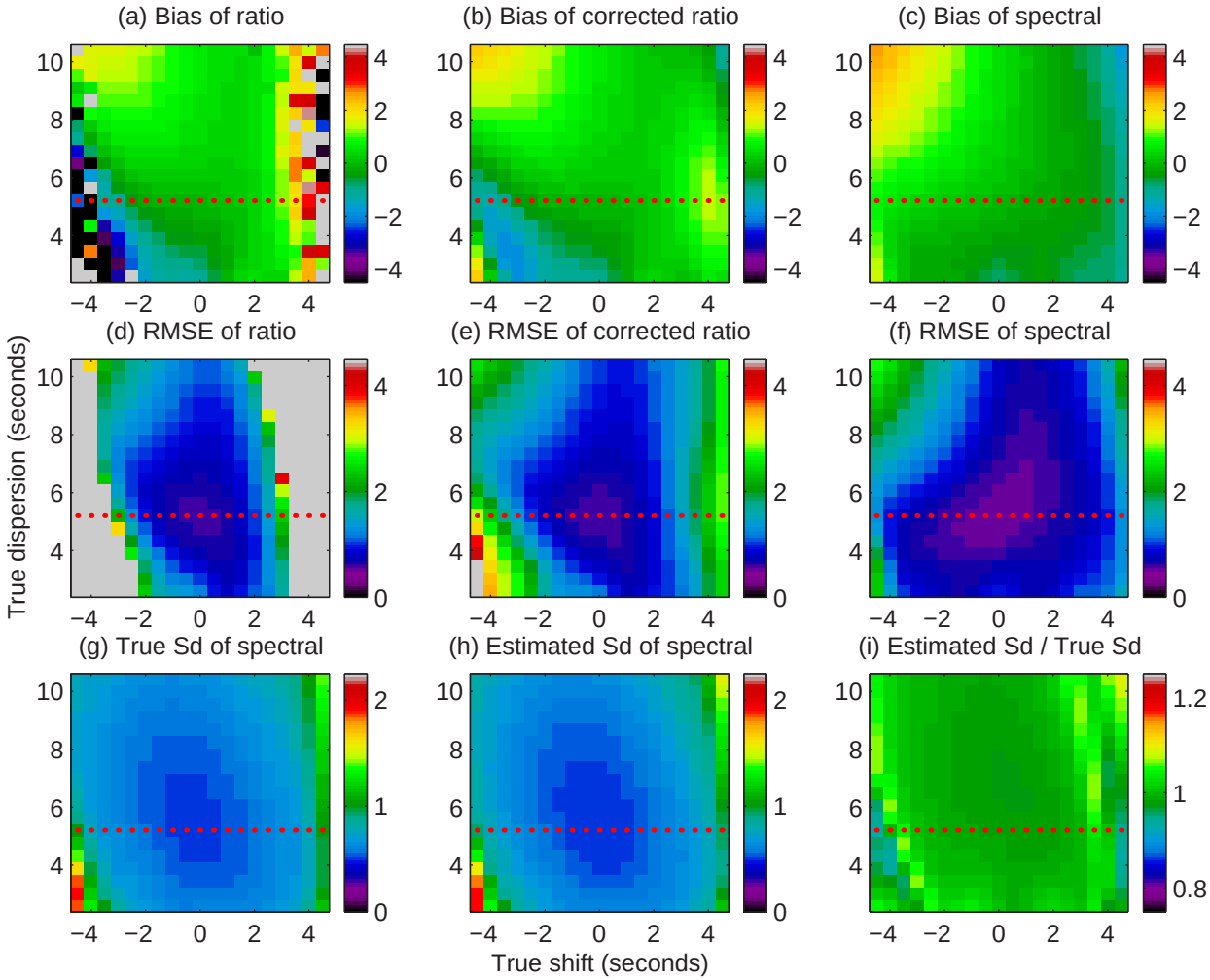

Figure 4: As for Figure 3, but varying the dispersion of the HRF for fixed standardised magnitude $\tau = 6$, indicated by a dotted horizontal line on Figure 3. The dotted horizontal line in this figure corresponds to the dispersion (5.2 seconds) used for Figure 3.

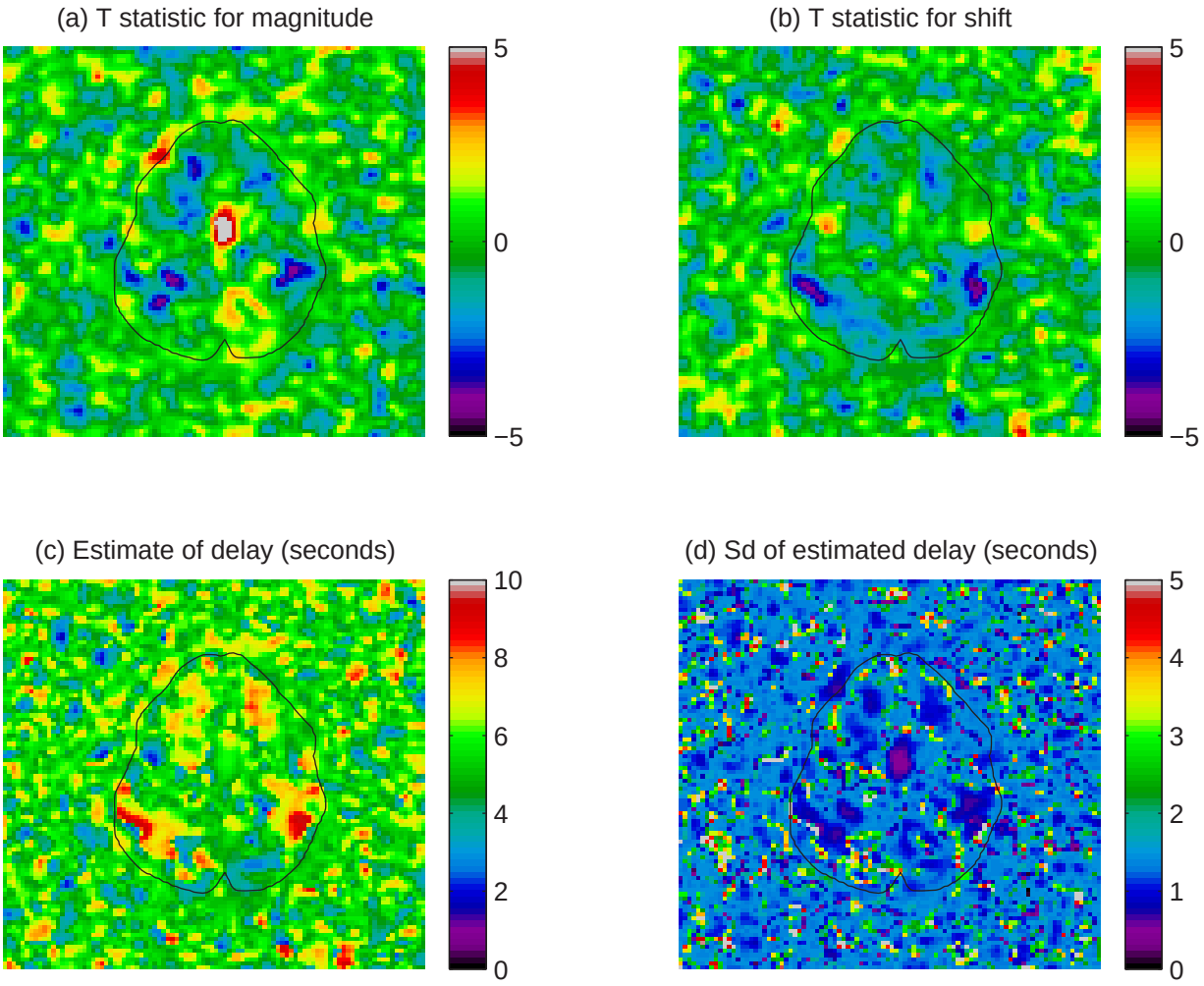

Figure 5: Inference for the delay of the pain stimulus with a reference delay of 5.4 seconds: (a) the T statistic for the magnitude, T_0, (b) the T statistic for the shift, T_1, (c) the estimated delay $\widehat{d}$ (seconds), (d) the standard deviation of the estimated delay $\widehat{\mathrm{Sd}}(\widehat{d})$ (seconds). Only one slice (slice 4) is shown.

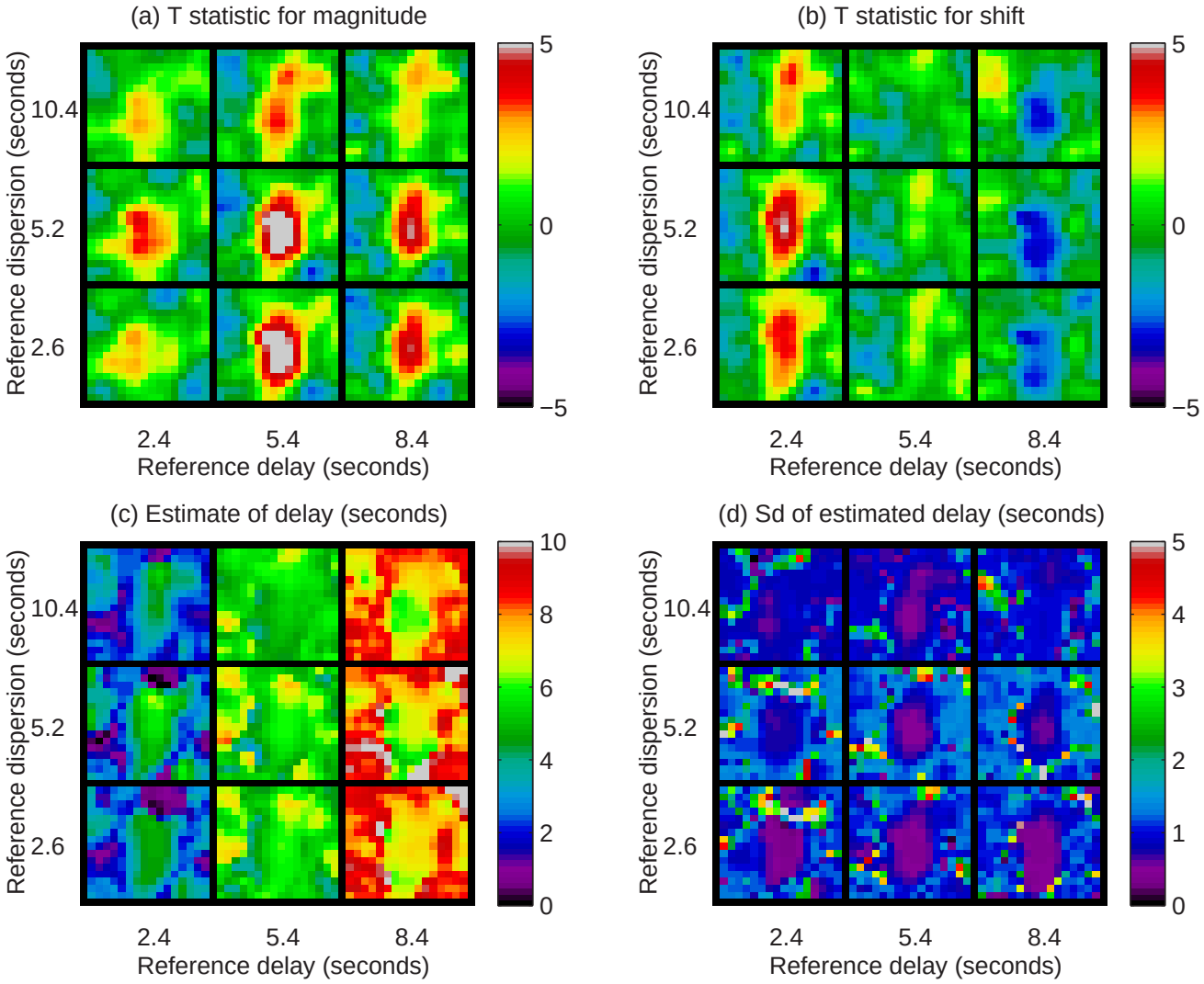

Figure 6: Magnification of the region of significant signal in Figure 5 (center of the 3×3 array) together with different reference delays and dispersions.

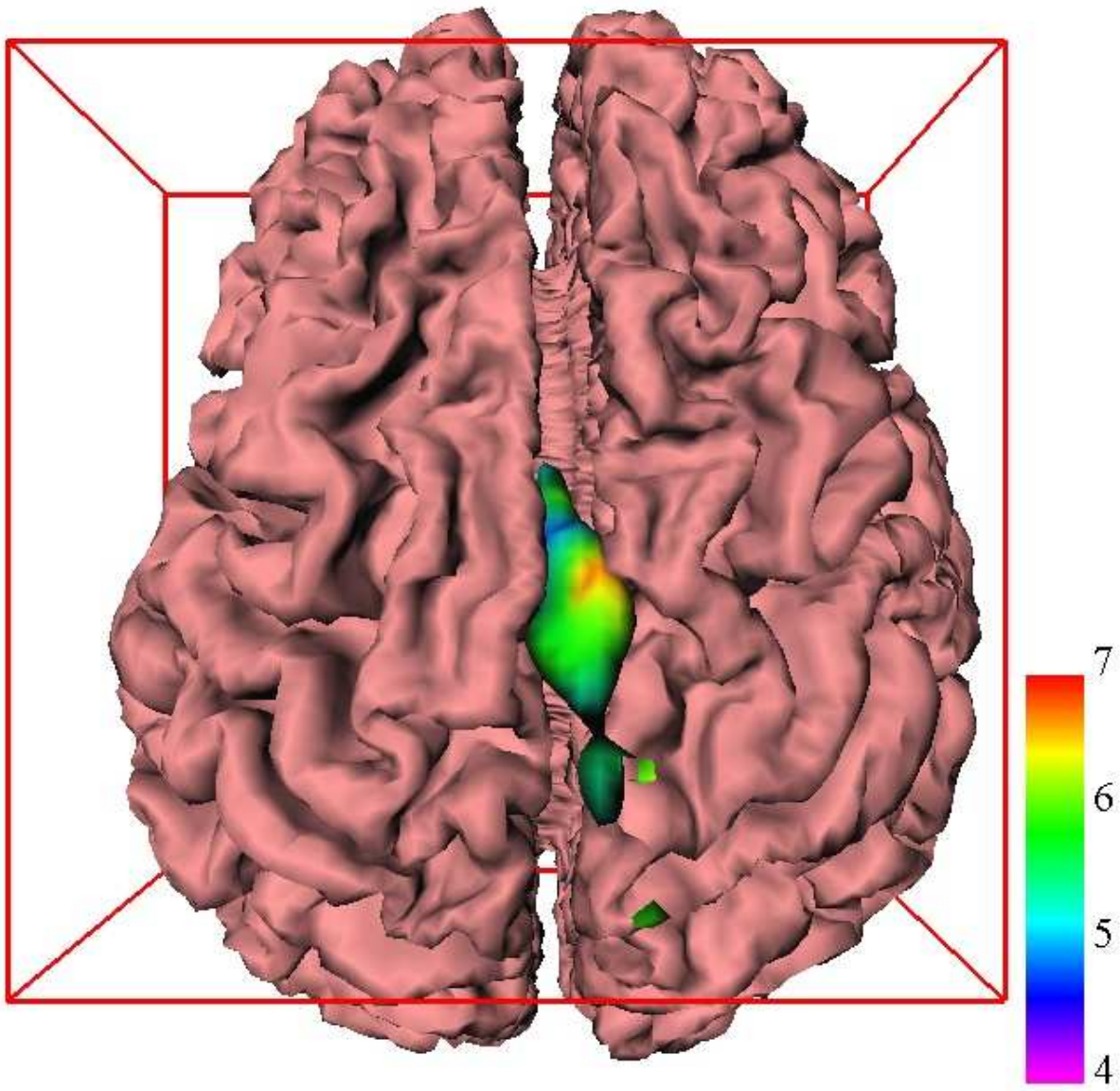

Figure 7: The T statistic T_0 for the magnitude of the hot stimulus thresholded at the $P = 0.05$ value of 4.86 - top view (coloured) together with the mid cortical surface (pink). The colours code for the estimated delay of the hemodynamic response, in seconds.

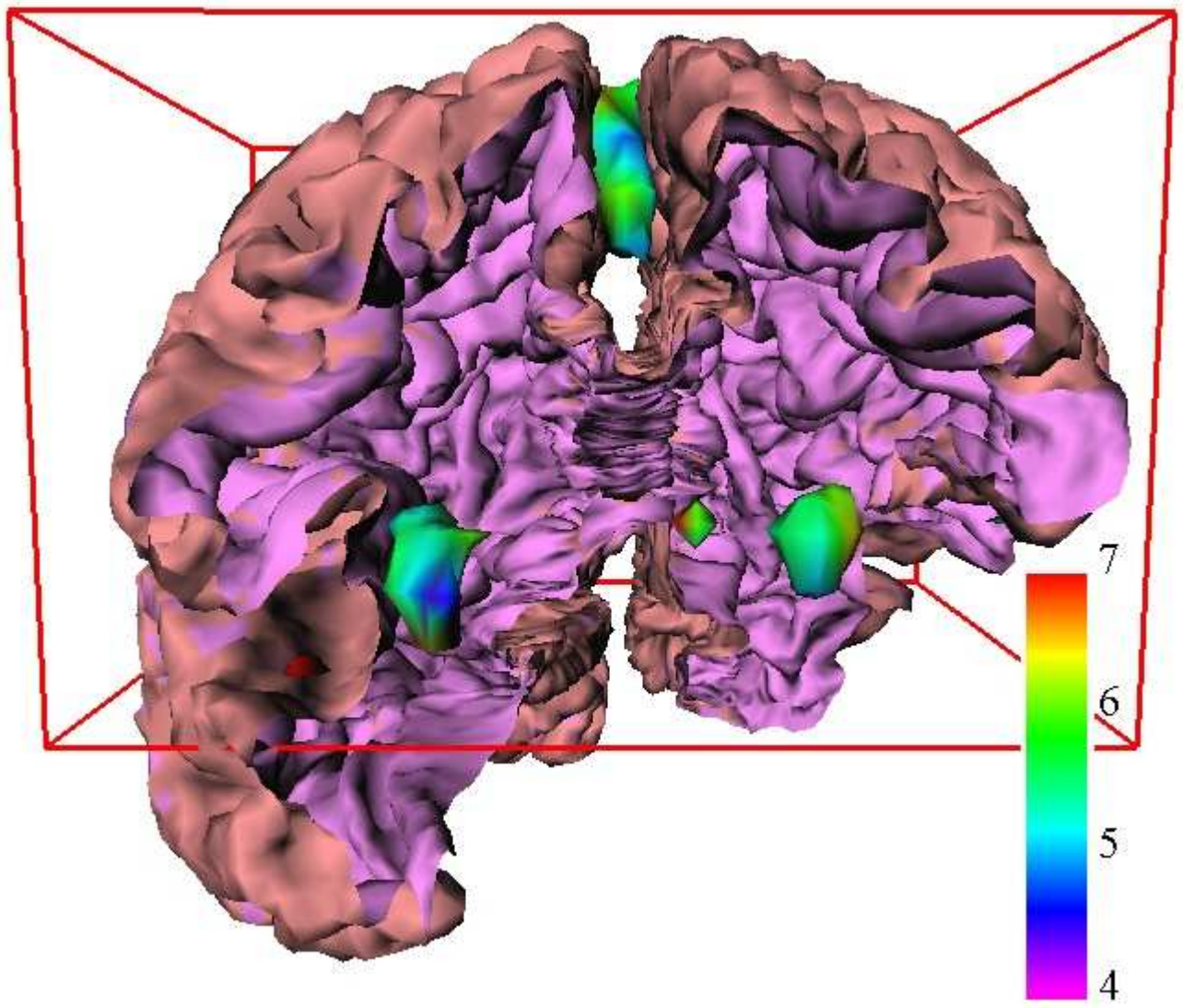

Figure 8: As for Figure 7: front view with part of the frontal cortex removed to show activation inside the brain.

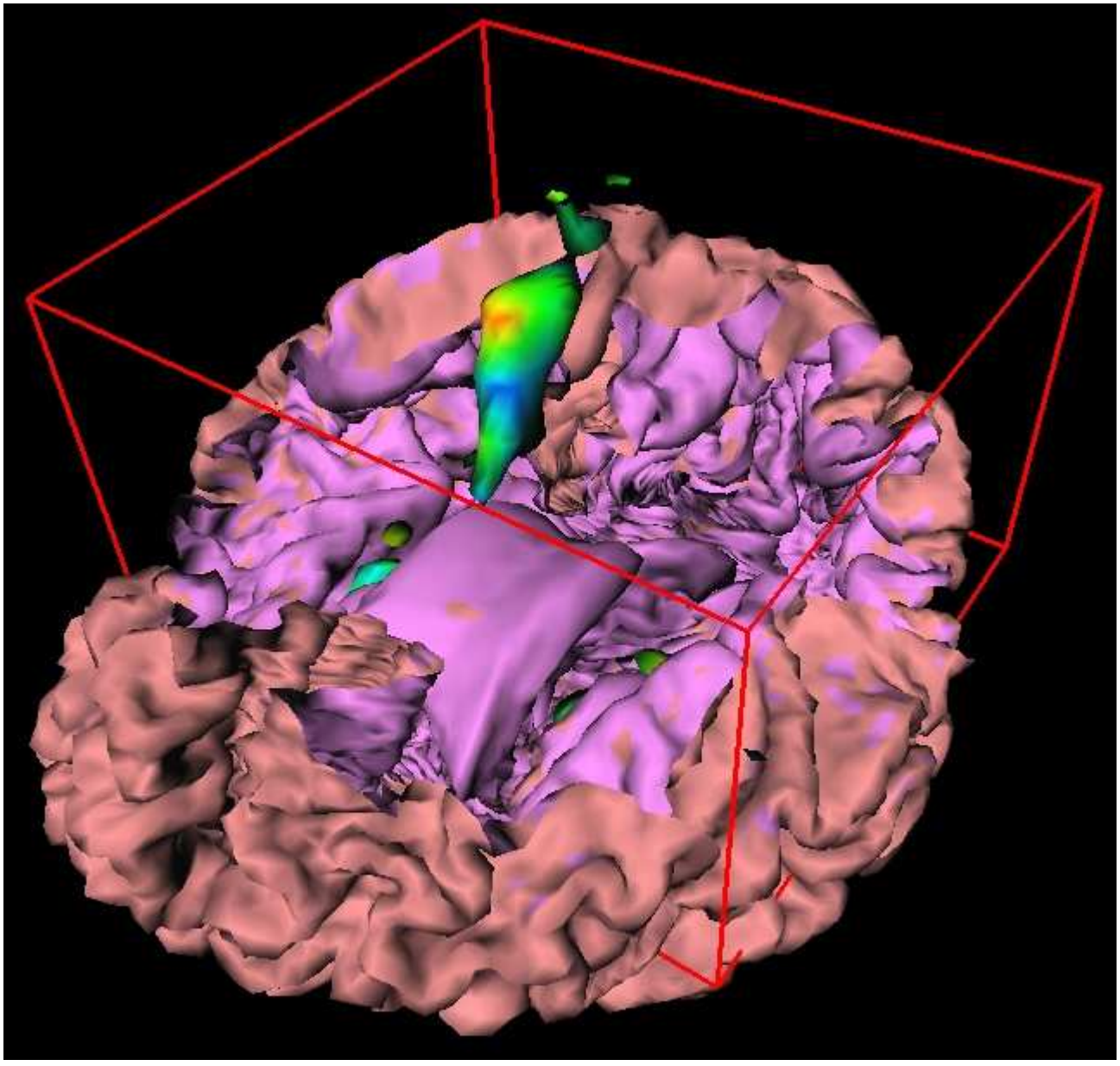

Figure 9: Possible cover for *NeuroImage*