Detecting fMRI activation allowing for unknown latency of the hemodynamic response

K.J. Worsley
McGill University

J.E. Taylor
Stanford University

January 7, 2006

Abstract

Several authors have suggested allowing for unknown latency of the hemodynamic response by incorporation of hemodynamic derivative terms into the linear model for the statistical analysis of fMRI data. In this paper we show how to use random field theory to provide a P-value for local maxima of two test statistics that have been recently proposed for detecting activation based on this analysis.

1 Introduction

The most common method for the statistical analysis of fMRI data uses the linear model that includes covariates given by the stimulus convolved with a known hemodynamic response function (hrf). There has been concern about the latency of the hrf, caused for example by small hemodynamic delays or slice-timing discrepancies. Friston et al. (1998) suggested allowing for unknown latency by adding the stimulus convolved with the derivative of the hrf to the linear model. The simplest case of fMRI data $Y(t)$ at time t, one stimulus $s(t)$, no temporal correlation, no drift, and hrf $h(t)$ with unknown delay or latency shift δ, is the model

$$Y(t) = (s \star h)(t - \delta)\beta + \epsilon(t) \approx x_1(t)\beta_1 + x_2(t)\beta_2 + \epsilon(t) \tag{1}$$

where $x_1 = (s \star h)$, $x_2 = -(s \star \dot{h})$, $\star$ is convolution, dot is derivative, $\beta_1 = \beta$, $\beta_2 = \beta\delta$ and $\epsilon(t) \sim \mathrm{N}(0, \sigma^2)$ independently. Henson *et al.* (2002) and Liao *et al.* (2002) present methods for estimating the delay within the context of this model. Our main interest here is detecting the activation allowing for unknown delay. The broader question of the relevance of this model is discussed more fully in Henson *et al.* (2002), Liao *et al.* (2002), Friman *et al.* (2003) and Calhoun *et al.* (2004).

We can detect activation by testing either β_1 (with a T statistic, T_1) or both β_1, β_2 (with an F statistic, F), as follows. Let $\widehat{\beta}_j$ denote the least-squares estimator of β_j, and let $\widehat{\sigma}_j$ denote its estimated standard deviation. The T statistic for β_j is

$$T_j = \widehat{\beta}_j / \widehat{\sigma}_j, \quad j = 1, 2, \tag{2}$$

which has a null T distribution with degrees of freedom ν equal to the number of temporal observations minus the number of regressors. We shall assume at first that x_1 and x_2 are orthogonal, which is true if their is a a sufficient period of rest at the beginning and end of the run (since $\int h\dot{h} = 0$). We shall relax this assumption later on in Section 4. Then $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are independent, so that the F statistic is

$$F = (T_1^2 + T_2^2)/2. \tag{3}$$

which has a null F distribution with $2, \nu$ degrees of freedom.

Neither choice is very satisfactory; the T statistic T_1 is the best choice if $h(t)$ matches the true hrf, but if the true hrf is different then we will lose sensitivity. The F statistic F is sensitive to *any* linear combination of the hrf and its derivative, but it will waste sensitivity on unrealistic linear combinations. In particular, it will waste sensitivity on negative BOLD responses to the stimulus. This is because the threshold for the F statistic must be set higher in order to avoid finding false positive activations over all (realistic and unrealistic) linear combinations. Because of the higher threshold, it may fail to detect true activations for a realistic linear combination, hence it may be less sensitive.

To concentrate the F statistic on positive BOLD responses, Calhoun *et al.* (2004) have proposed a 'one-sided' F statistic that takes the same value as F but multiplied by the sign of $\widehat{\beta}_1$:

$$F' = \mathrm{sign}(T_1)(T_1^2 + T_2^2)/2. \tag{4}$$

This should be more sensitive than the (two-sided) F statistic F at detecting positive BOLD responses.

Friman *et al.* (2003) have proposed something more subtle. He noted that δ is usually confined to a known range of plausible values $[\Delta_1, \Delta_2]$, such as $[-2, 2]$ seconds. The set of values of β_1, β_2 is then confined to a *cone* centred at the origin, known as a *cone alternative* (Lin & Lindsay, 1997; Takemura & Kuriki, 1997) or *cone classes* (Ramprashad *et al.*, 1996). The likelihood ratio test statistic (under known variance σ^2) is then the maximum T statistic for testing the regressor $(x_1(t) + x_2(t)\delta$ for fixed δ, maximized over all values of δ in $[\Delta_1, \Delta_2]$. We call this the *cone T statistic*:

$$T' = \max_{\delta \in [\Delta_1, \Delta_2]} \frac{T_1 \widehat{\sigma}_1 + T_2 \widehat{\sigma}_2 \delta}{\sqrt{\widehat{\sigma}_1^2 + \widehat{\sigma}_2^2 \delta^2}}. \tag{5}$$

This can be expressed in angles as follows. Let $\theta = \arctan(\delta\widehat{\sigma}_2/\widehat{\sigma}_1)$, $L_j = \Delta_j \widehat{\sigma}_2/\widehat{\sigma}_1$, and $\Theta_j = \arctan(L_j)$, $j = 1, 2$. Then $\Theta_2 - \Theta_1$ is the angle of the cone and

$$T' = \max_{\theta \in [\Theta_1, \Theta_2]} T_1 \cos\theta + T_2 \sin\theta. \tag{6}$$

A more convenient form for computations is

$$T' = \max\left\{ \frac{T_1 + T_2 L_1}{\sqrt{1 + L_1^2}}, \frac{T_1 + T_2 L_2}{\sqrt{1 + L_2^2}}, (T_1 > 0)(L_1 \leq T_2/T_1 \leq L_2)\sqrt{T_1^2 + T_2^2} \right\}. \tag{7}$$

The four tests we have considered can all be compared in terms of their rejection regions, the set of values of T_1, T_2 that would cause us to reject the null hypothesis of no activation. These are plotted in Figure 1.

The purpose of this paper is to use random field theory to find an approximate P-value for local maxima of SPMs of the Calhoun one-sided F statistic F' and the Friman cone T statistic T'. We shall then compare the sensitivity of these two tests to the simple T statistic T_1 for the hrf, and the F statistic F for the hrf and its derivative.

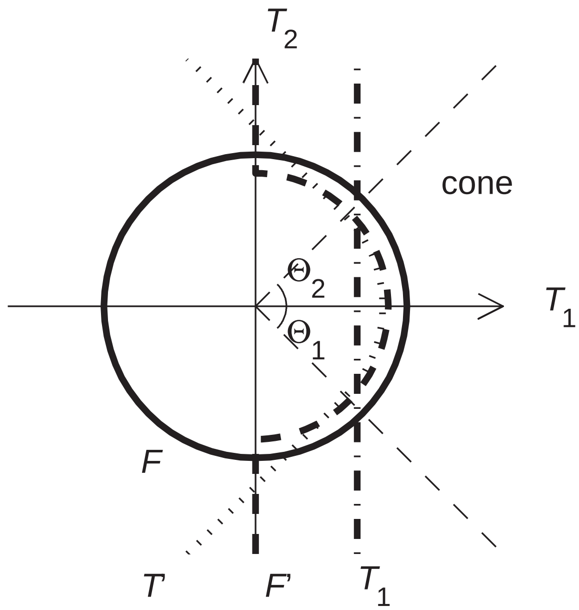

Figure 1: Rejection regions (the side of the boundary that excludes the origin) of the four tests at $P = 0.05$ (cone limits are $\Theta_1 = -45^o$, $\Theta_2 = 45^o$).

2 Specificity

2.1 T statistic T_1

For completeness, we give the approximate P-value of local maxima of T_1 inside a search region S in D dimensions, based on the expected Euler characteristic (EC):

$$\mathrm{P}(\max_S T_1 \geq t) \approx \sum_{d=0}^{D} Resels_d(S)\ EC_d(t) \tag{8}$$

where $Resels_d(S)$ is the d-dimensional resels of the search region S, and $EC_d(t)$ is the EC density of the T statistic SPM with ν degrees of freedom (Worsley *et al.*, 1996a).

2.2 F statistic F

First, as suggested by Calhoun *et al.* (2004), let us consider the more general situation where we have p basis functions instead of two, and p T statistics $T_1, \ldots, T_p$ whose numerators are independent. Then define the more general F statistic by

$$F = (T_1^2 + \cdots + T_p^2)/p. \tag{9}$$

Again for completeness, the approximate P-value of local maxima of F is

$$\mathrm{P}(\max_S F \geq t) \approx \sum_{d=0}^{D} Resels_d(S)\ EC_d^p(t) \tag{10}$$

where $EC_d^p(t)$ is the EC density of the F statistic SPM with p, ν degrees of freedom (Worsley *et al.*, 1996a).

2.3 One-sided F statistic F'

Since the rejection region of the one-sided F statistic is exactly half that of the usual F statistic (see Figure 1), then at a fixed voxel, the P-value of the Calhoun test is exactly half the P-value of the F test. However this does not quite hold true for local maxima of the corresponding SPMs. The reason is a subtle one, and has to do with the fact that local maxima can occur on the boundary where $\text{sign}(T_1)$ changes from +1 to -1, as well as the interior where $\text{sign}(T_1) = 1$.

Calhoun *et al.* (2004) considered the more general one-sided F statistic

$$F' = \text{sign}(T_1)(T_1^2 + \cdots + T_p^2)/p. \tag{11}$$

The approximate P-value of local maxima of F', based on the expected EC, is given by

$$\text{P}(\max_S F' \geq t) \approx \tfrac{1}{2}\sum_{d=0}^{D} \mathit{Resels}_d(S)\ EC_d^p(t) + \sum_{d=0}^{D} \mathit{Resels}_d^0(S)\ EC_d^{p-1}(tp/(p-1)) \tag{12}$$

where $\mathit{Resels}_d^0(S)$ is the expected resels of the boundary where $T_1 = 0$, given by

$$\mathit{Resels}_d^0(S) = \sum_{i=0}^{\lfloor (D-d-1)/2 \rfloor} \frac{(-1)^i (4\log 2)^{i+\frac{1}{2}} \Gamma\left(i+\frac{1}{2}\right) \Gamma\left(\frac{d}{2}+i+1\right)}{2\pi^{i+1}\Gamma(i+1)\Gamma\left(\frac{d+1}{2}\right)} \mathit{Resels}_{d+2i+1}(S)$$

which is of course zero at $d = D$ (Taylor & Worsley, 2005). For those curious about the derivation of this result, the second term in (12) is a conjunction of a unit Gaussian SPM at 0 and an F statistic SPM with $p-1, \nu$ degrees of freedom above $tp/(p-1)$ (Worsley & Friston, 2000; Friston *et al.*, 1999). In practice the second term adds very little (≈ 0.01) to the $P = 0.05$ threshold for typical whole-brain searches.

2.4 Cone T statistic T'

The random field P-value for the cone T statistic T' can be determined by adding an extra dimension, the 'cone' dimension, to the search region. In other words, the search region becomes the $D+1$ dimensional space of all spatial locations in S and all cone angles θ in the interval $[\Theta_1, \Theta_2]$. This is analogous to 'scale space' (Worsley *at al.*, 1996b) where an extra parameter is added for the scale of the smoothing filter.

All that is required is to compute the resels of this new 'cone space'. In the cone angle (in radians), the effective FWHM is simply $(4\log 2)^{1/2}$, and so the resels of the cone C are $\mathit{Resels}_0(C) = 1$ and $\mathit{Resels}_1(C) = (4\log 2)^{-1/2}(\Theta_2 - \Theta_1)$. Taylor & Worsley (2005) show that an approximate P-value is

$$\text{P}(\max_S T' \geq t) \approx \sum_{d=0}^{D} \mathit{Resels}_d(S) \sum_{k=0}^{1} \mathit{Resels}_k(C)\ EC_{d+k}(t). \tag{13}$$

Note that this approximation is the expected EC only if the residual degrees of freedom ν is infinite.

3 Sensitivity

First of all there is an interesting theoretical result that says that the one-sided F statistic F' is *inadmissable*. This is a concept from decision theory; it means that there exists another test with the same P-value that is more sensitive for *any* alternative. This follows from a result of Birnbaum (1954) which states that any test with a concave acceptance region (the opposite of the rejection region) is inadmissable. As we can see from Figure 1, the acceptance region of the one-sided F statistic is concave, so it is inadmissable. The other three test statistics have convex acceptance regions, so they are *admissable*, meaning that there is no other test that is *always* more sensitive. In passing, note that the minimum test for a conjunction (Friston *et al.*, 1999; Worsley & Friston, 2000; Wilkinson's Case 2 from Brinbaum, 1954) is also inadmissable because it has a concave acceptance region. Unfortunately Birnbaum's proof merely says that a more sensitive test must exist, but gives no method for constructing it. In particular it is not clear which if any of the three other tests is better than the one-sided F statistic. Moreover the Birnbaum result only applies to the test at a single voxel, and not to local maxima over a search region.

To give some idea of sensitivity, we compared the four test statistics under 'ideal' conditions. That is, we assumed the search region was a 3D ball with volume 1000cc and a FWHM of 10mm, and that the residual degrees of freedom was infinite ($\nu = \infty$). We assumed that we have $p = 2$ basis functions. This establishes the $P = 0.05$ thresholds for the first three test statistics, which are, on the same scale of minimum distance of the rejection region from the origin, $T_1 = 4.66$, $\sqrt{2F} = 5.21$ and $\sqrt{2F'} = 5.07$.

The fourth test statistic, the cone T statistic, depends on the angle of the cone, which in turn depends on the hrf, its derivative, the stimulus, and the range of latencies. For the hrf we chose the sum of two gamma functions suggested by Glover (1999), the first with a latency of 5.4s, a spread of 5.2s and a weight of 1, the second with a latency of 10.8s, a spread of 7.35s and a weight of -0.35. We chose two stimuli: an event-related stimulus, and a 20s block stimulus. The covariates x_1 and their derivatives x_2 are shown in Figure 2. We assume that there is a long period of rest before and after the stimulus so that x_1 and x_2 are orthogonal and $\widehat{\sigma}_j^2 \propto 1/\int x_j^2$. We chose a range for the unknown latency shift of $[\Delta_1, \Delta_2] = [-2, 2]$s. This gives a cone angle of $\Theta_2 - \Theta_1 = 78.4^o$ for the event-related design, and 38.1^o for the block design. These in turn gave $P = 0.05$ thresholds of 4.95 and 4.84, respectively. The reason that the block design threshold is lower is that the cone is smaller and so its resels are smaller.

To get a rough estimate of sensitivity, we also supposed that there was sufficient activation at a single voxel to ensure that the maximum of the SPM is always located at that voxel (Siegmund & Worsley (1995) show that asymptotically this gives the correct sensitivity). The sensitivity of each test is then approximated by integrating the bivariate normal density for T_1, T_2 inside its rejection region. The standard deviation of each T_j is 1, but its mean depends on the magnitude β of the stimulus and the latency δ of the hrf. This is

$$\mathrm{E}(T_j) = \mu \frac{\int x_1(t-\delta) x_j(t) dt}{\sqrt{\int x_1(t)^2 dt \int x_j(t)^2 dt}}$$

where μ depends on β/σ, but not on δ or j, $j = 1, 2$. We chose $\mu = 5.5$, corresponding to $\mathrm{E}(T_1) = 5.5$ when the hrf is correct ($\delta = 0$). The resulting sensitivities, plotted as a function of the shift of the hrf δ, are shown in Figure 2. Note that the plot is symmetric in δ, so only positive values of δ are shown.

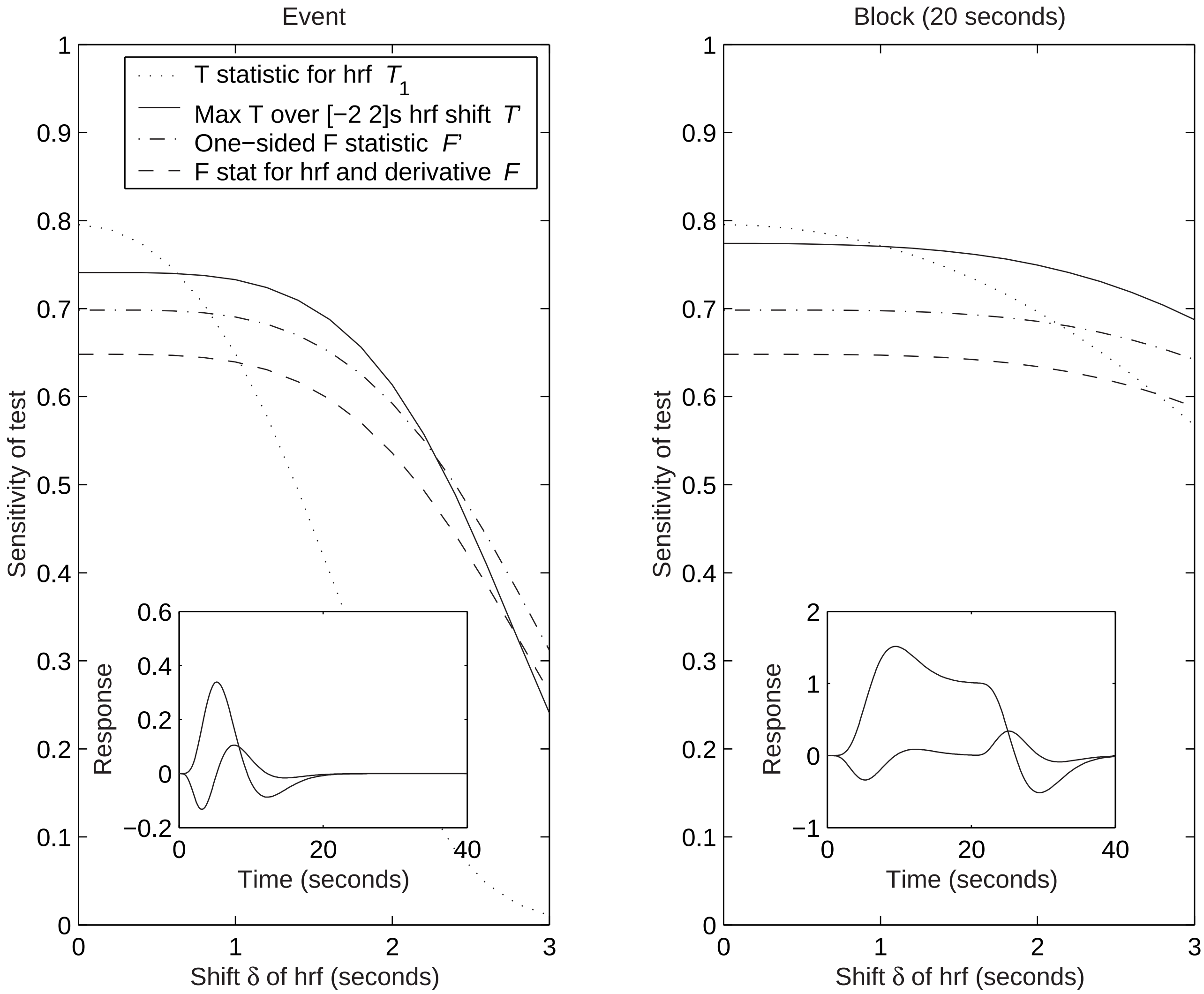

Figure 2: Sensitivities of the of the four tests at $P = 0.05$ as a function of the shift δ of the hrf for an event related design (left) and a block design (right). The responses and their derivatives are shown in the lower panels.

For shifts less than $\sim$1s, the simple T statistic is the most sensitive, and for larger shifts the cone T statistic is the most sensitive. For very large shifts the one-sided F statistic is the most sensitive. The simple F statistic is always less sensitive. The same overall pattern is repeated for the block design, though the sensitivities do not vary as greatly with shift in the hrf. Our overall recommendation is to use the simple T statistic if the shift in the hrf is known to be less than 1s, but if the shift in the hrf is uncertain, then the cone T statistic seems to be the best choice.

4 Application

In practice the linear model is a lot more complicated than the simple model (1). There are usually many other regressors for drift and for other stimuli, and the data is temporally correlated. This introduces a correlation c between $\widehat{\beta}_1$ and $\widehat{\beta}_2$ which were assumed independent so far. The above methods can still be applied provided we replace T_2 by $(T_2 - cT_1)/\sqrt{1-c^2}$ in the F statistics (3) and (4). We cannot use (5) for defining the cone T statistic, but we can still use (7) provided we change the limits. To see how to do this, note that conditional on the variance estimates,

$$\begin{aligned} \mathrm{E}(T_1) &= \beta/\widehat{\sigma}_1, \\ \mathrm{E}(T_2) &= \frac{\beta\delta/\widehat{\sigma}_2 - c\beta/\widehat{\sigma}_1}{\sqrt{1-c^2}}, \end{aligned}$$

which is still a cone alternative but with limits L_j replaced by $(L_j - c)/\sqrt{1-c^2}$, $j = 1, 2$. There is now a problem with the P-value approximation (13): since temporal correlation varies spatially, then c varies spatially and so the cone angle $\Theta_2 - \Theta_1$ (which determines the P-value approximation) also varies spatially. However we shall see that in practice $\Theta_2 - \Theta_1$ is remarkably constant, so we suggest replacing it by its average value in the P-value approximation (13).

We applied the above test statistics to detecting a hot stimulus in a pain perception experiment fully described in Worsley *et al.* (2002). The paradigm was 9s alternating hot and warm stimuli applied to the right calf, interspersed with 9s rest, repeated 10 times (1.5T, TR=3s). A block design was used with the same hrf as in Section 3. A cubic polynomial in the frame number was added to account for drift. The errors were assumed to follow an AR(1) model, which was estimated and smoothed by the methods given in Worsley *et al.* (2005), to give an estimated $\nu = 97$ degrees of freedom.

For a whole brain search with volume 1184cc and an estimated 8.78 mm FWHM average smoothing, the $P = 0.05$ for the first three test statistics, on the same scale of minimum distance of the rejection region from the origin, are $T_1 = 5.15$, $\sqrt{2F} = 5.80$ and $\sqrt{2F'} = 5.63$. For a $[-2, 2]$s hrf shift range, the cone angle $\Theta_2 - \Theta_1$ was remarkably constant, averaging at 60.9 ± 1.7^o, giving a threshold of $T' = 5.46$.

Figure 3 shows the four thresholded images. The volume of detected activation due to the hot stimulus remains roughly the same: T_1 : 4.0cc, F : 2.9cc, F' : 3.8cc and T' : 4.3cc. Interestingly, it is the cone T statistic with shifts in the range $[-2, 2]$s that detects the most activation (left primary somatosensory area and left and right thalamus).

Finally it must be admitted that due to the large voxel size (2.3×2.3×7.0mm), Bonferroni thresholds were actually lower. These can be calculated for each of the four test statistics by setting the resels of the search region to $[1, 0, 0, 0]$ and multiplying the resulting P-values by the number of voxels in the search region. Since our objective is to illustrate and compare methods, we do not report these results here.

For an image of latencies and their sd for this data see Liao *et al.* (2002), Figure 5. For an analysis on more subjects see Henson *et al.* (2002) and Calhoun *et al.* (2004).

5 Conclusions

We have presented methods for calculating the specificity of two new test statistics for detecting activation allowing for unknown latency in the hrf. The tests are based on approximating the hrf by a canonical hrf and its derivative. The first test, proposed by Calhoun *et al.* (2004), multiplies the usual F statistic by the sign of the T statistic for the canonical hrf. The second, proposed by Friman *et al.* (2003), is based on maximising the T statistic over

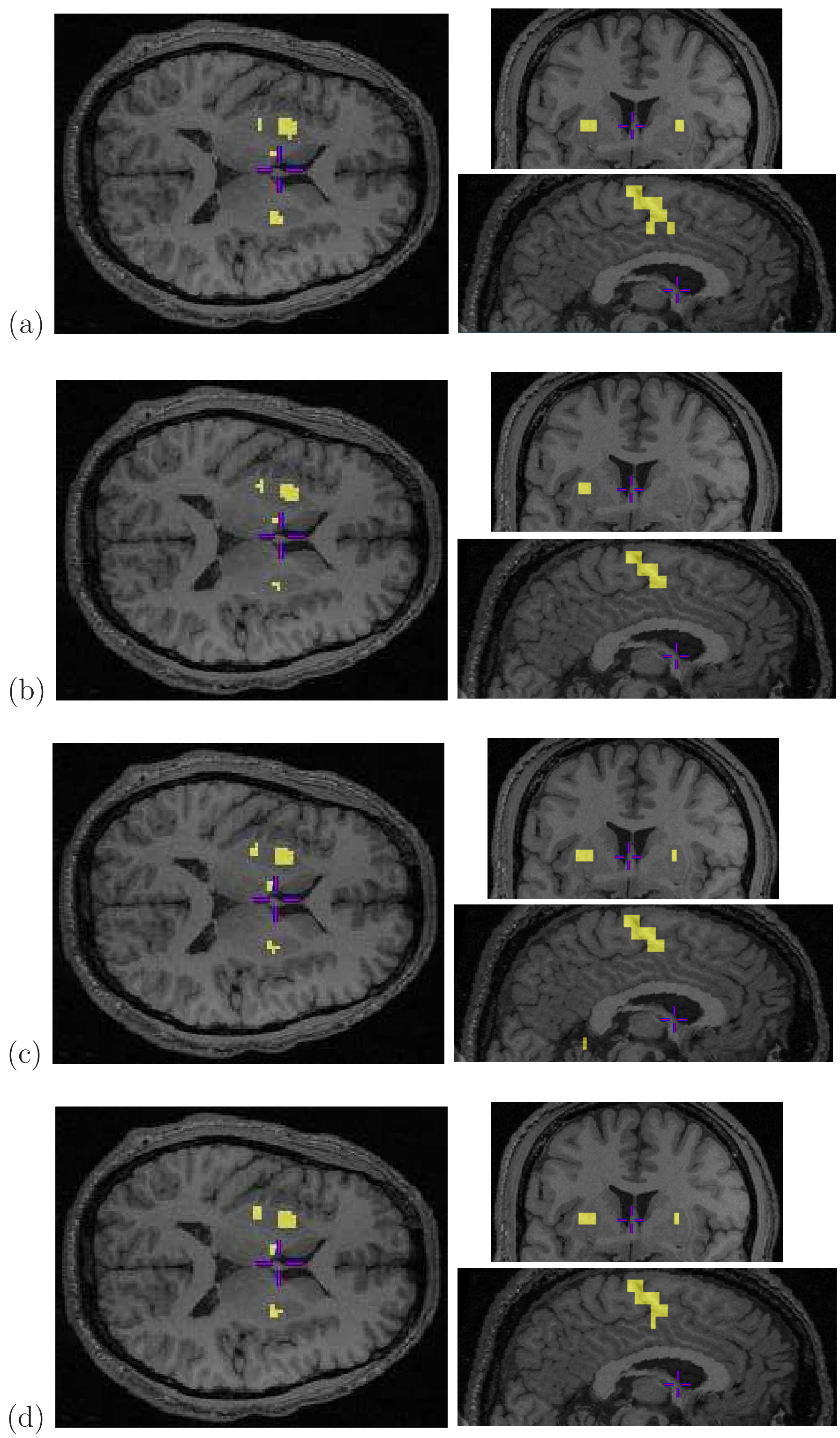

Figure 3: Hot stimulus detected at $P = 0.05$ by (a) T statistic T_1: 4.0cc, (b) F statistic $\sqrt{2F}$: 2.9cc, (c) one-sided F statistic $\sqrt{2F'}$: 3.8cc, (d) cone T statistic T': 4.3cc. The cone T statistic detects the most activation.

a range of shifts, known as a cone T statistic. Our main results are P-value approximations for local maxima based on random field theory.

We then compared the sensitivity of these two tests with two others that have been proposed some time ago: the simple T statistic for the canonical hrf, and the usual F statistic for the canonical hrf and its derivative. Overall the simple T statistic is most sensitive when the shift in the hrf is ∼1s, but for larger shifts the most sensitive test appears to be the cone T statistic. This conclusion is the same for both event-related and block designs, although for the latter the sensitivities are less dependent on the shift in the hrf, as we would expect.

A small point to consider is whether to include the temporal derivative at all if we know that the shift in the hrf is less than 1s. It should make little difference: the derivative is orthogonal to the hrf (at least if no other terms are present in the model) so adding or removing it should not alter parameter estimates or inference about the activation. The only price to pay for adding the derivative term is the loss of one degree of freedom, negligible compared the residual degrees of freedom which is usually at least 100. For the application in Section 4, the maximum T statistic with and without the derivative term was identical to two decimals (7.42).

We then applied our results to an experiment in pain perception. Despite differences in the $P = 0.05$ thresholds for the four tests (e.g. the cone T statistic threshold must be larger than the simple T statistic threshold), the volume of detected activation was roughly the same. Nevertheless it was the cone T statistic that detected the most activation, perhaps because of uncertainty in the shift of the hrf.

These methods were all applied to simple stimulus detection. How can they be applied to *contrasts* in stimuli, such as a difference between the hot and warm stimuli in our example? In principle the same methods can be used, provided we assume that the unknown shift in the hrf is the *same for all stimuli* at a given voxel. Then we can replace T_1 by the T statistic for the contrast using the canonical hrf, and T_2 by the T statistic for the contrast using the derivative of the hrf. To generalise to the more interesting case of *different shifts for different stimuli* at the same voxel would require pairs of T statistics for each contrasted stimulus. This is beyond the scope of this paper.

We have not addressed the issue of how to combine these results across runs on the same subject, or across subjects in a population. In principle it is straightforward - the estimated effects $\widehat{\beta}_j$, their estimated standard deviations $\widehat{\sigma}_j$ and their correlation c could be combined an a mixed effects model at the next level in the hierarchy. The techniques presented here would then be used at the final stage in the hierarchy to obtain an overall test for activation. This will be the subject of further work.

References

Birnbaum, A. (1954). Combining independent tests of significance. *Journal of the American Statistical Society*, **49**:559-574.

Calhoun, V.D., Stevens, M.C., Pearlson, G.D. & Kiehl, K.A. (2004). fMRI analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. *Neuroimage*, **22**:252-257.

Friman, O., Borga, M., Lundberg, P., & Knutsson, H. (2003). Adaptive analysis of fMRI data. *NeuroImage*, **19**:837-845.

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.

Friston, K.J., Holmes, A.P., Price, C.J., Büchel, C. & Worsley, K.J. (1999). Multi-subject fMRI studies and conjunction analyses. *NeuroImage*, **10**:385-396.

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

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 nonwords and initial versus repeated face presentations. *NeuroImage*, **15**:83-97.

Liao, C., Worsley, K.J., Poline, J-B., Duncan, G.H. & Evans, A.C. (2002). Estimating the delay of the response in fMRI data. *NeuroImage*, **16**:593-606.

Lin, Y. & Lindsay, B.G. (1997). Projections on cones, chi-bar squared distributions, and Weyl's formula. *Statistics and Probability Letters*, **32**:367-376.

Ramprashad, S., Parks, T.W. & Shenoy, R. (1996). Signal modelling and detection using cone classes. *IEEE Transactions on Signal Processing*, **44**:329-338.

Siegmund, D.O & Worsley, K.J. (1995). Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Annals of Statistics*, **23**:608-639.

Takemura, A. & Kuriki, S. (1997). Weights of $\bar{\chi}^2$ distribution for smooth or piecewise smooth cone alternatives. *Annals of Statistics*, **25**:2368-2387.

Taylor, J.E. & Worsley, K.J. (2005). Random fields of multivariate test statistics, with applications to shape analysis and fMRI. `http://www.math.mcgill.ca/keith/roy/roy.pdf`, submitted.

Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J. & Evans, A.C. (1996a). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, **4**:58-73.

Worsley, K.J., Marrett, S., Neelin, P. & Evans, A.C. (1996b). Searching scale space for activation in PET images. *Human Brain Mapping*, **4**:74-90.

Worsley, K.J. & Friston, K.J. (2000). A test for a conjunction. *Statistics and Probability Letters*, **47:**135-140.

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

Worsley, K.J. (2005). Spatial smoothing of autocorrelations to control the degrees of freedom in fMRI analysis. *NeuroImage*, in press.