NeuroImage **13,** 218–229 (2001)
doi:10.1006/nimg.2000.0711, available online at http://www.idealibrary.com on IDEAL®

Separating Processes within a Trial in Event-Related Functional MRI

II. Analysis

J. M. Ollinger,*,†,[1] M. Corbetta,*,† and G. L. Shulman*,†

**Department of Radiology and †Department of Neurology, Washington University, St. Louis, Missouri 63110*

Received April 17, 2000

Many cognitive processes occur on time scales that can significantly affect the shape of the blood oxygenation level-dependent (BOLD) response in event-related functional MRI. This shape can be estimated from event related designs, even if these processes occur in a fixed temporal sequence (J. M. Ollinger, G. L. Shulman, and M. Corbetta. 2001. *NeuroImage* 13: 210–217). Several important considerations come into play when interpreting these time courses. First, in single subjects, correlations among neighboring time points give the noise a smooth appearance that can be confused with changes in the BOLD response. Second, the variance and degree of correlation among estimated time courses are strongly influenced by the timing of the experimental design. Simulations show that optimal results are obtained if the intertrial intervals are as short as possible, if they follow an exponential distribution with at least three distinct values, and if 40% of the trials are partial trials. These results are not particularly sensitive to the fraction of partial trials, so accurate estimation of time courses can be obtained with lower percentages of partial trials (20–25%). Third, statistical maps can be formed from *F* statistics computed with the extra sum of square principle or by *t* statistics computed from the cross-correlation of the time courses with a model for the hemodynamic response. The latter method relies on an accurate model for the hemodynamic response. The most robust model among those tested was a single gamma function. Finally, the power spectrum of the measured BOLD signal in rapid event-related paradigms is similar to that of the noise. Nevertheless, high-pass filtering is desirable if the appropriate model for the hemodynamic response is used.

INTRODUCTION

A major goal of studies using blood oxygenation level-dependent (BOLD) weighted functional MRI (FMRI) (Ogawa *et al.,* 1990; Kwong *et al.,* 1992) has been to identify regions of the brain that are activated while performing specific cognitive tasks. The most widely used approaches to this problem begin by computing a map of *t* statistics that test at each voxel the null hypothesis of no activation (Friston *et al.,* 1995a,b; Worsley *et al.,* 1995). A threshold is then applied to isolate statistically significant voxels or regions (Worsley *et al.,* 1992, 1995; Friston *et al.,* 1994). Computation of these statistical maps requires two quantities: the magnitude of the response and its variance at each voxel. These parameters are usually estimated from linear models of the data (Friston *et al.,* 1995a,b). These models are usually constructed by postulating a functional form for the BOLD response and using this function as a regressor in the model. The BOLD response is easy to characterize for block designs because their long task and control periods can be modeled by a simple square wave. Modeling the hemodynamic response as a linear system improves this model by more accurately modeling the rise time, delay, and fall time of the BOLD response (Boynton *et al.,* 1996).

Event-related designs (Buckner *et al.,* 1996) are more difficult to analyze because the task period is short with respect to the hemodynamic response. Therefore, the shape of the BOLD response is dominated by the shape of the hemodynamic response, which is known to vary across subjects and regions of the brain (Lee *et al.,* 1995; Buckner *et al.,* 1996, 1998; Kim *et al.,* 1997; Schacter *et al.,* 1997; Aguirre *et al.,* 1998). Moreover, the shape of the responses may depend on the experimental paradigm. For example, the BOLD response in areas that maintain information over a delay period (such as in match-to-sample tasks)

[1] To whom correspondence and reprint requests should be addressed at Washington University School of Medicine, Neuroimaging Laboratory, Campus Box 8225, St. Louis, MO 63110. Fax: 314-362-6110. E-mail: jmo@npg.wustl.edu.

1053-8119/01 $35.00

will depend on the duration of the interval. The shape of the hemodynamic response therefore carries important information about the cognitive processes related to the observed activation.

In the companion paper (Ollinger *et al.*, 2001), we introduced and validated a method for separately estimating the time course of two successive BOLD responses occurring at a fixed interval. This method makes no assumptions about the shape of the hemodynamic response. It only requires that the sequence of trials during the experiment includes partial trials, in which only the first BOLD response is present, randomly intermixed with trials in which both BOLD responses are present. Moreover, the interval between trials (the intertrial interval or ITI) must be randomly varied. Here we consider four issues related to the analysis of these time course data:

1. What is the effect of correlations among points in the estimated time courses?
2. What are the optimal values of parameters of the experimental design?
3. How can the time courses derived from this method be statistically analyzed?
4. How is the estimation and statistical analysis of these time courses affected by high-pass filtering the data?

Effect of correlations. Individual points in time courses estimated with the linear model will in general be correlated. These correlations result primarily from the overlap of nearby responses in rapidly presented event-related studies. We discuss how these correlations affect the interpretability of the data.

Optimal experimental design. Optimal designs are those that yield the largest estimated magnitudes with the best statistical properties while satisfying the behavioral constraints of the experiment. We define the best statistical properties as low variance of the estimated effects, equal variance across effects, and minimum correlation among effects. We present the results of Monte Carlo simulations used to determine the distribution of ITIs and percentage of partial trials that yield the best statistical properties.

Statistical analysis of time courses. Statistical maps must be generated from the time courses in order to identify activated regions. We evaluate several different methods for computing these maps. The first method computes an F statistic using the extra sum-of-squares approach (Beck and Arnold, 1977). An interesting aspect of this technique is that significance levels are determined without making any assumption about the shape of the hemodynamic response. Statistical maps can also be generated by cross-correlating the estimated time courses with a model for the hemodynamic response and then forming t statistics. Since the resulting magnitude is strongly affected by the shape and timing of the model function, variations in the shape of the observed BOLD response with condition can produce spurious changes in magnitude. Therefore, it is important that the models be as robust as possible. In this paper, we evaluate the robustness of two models that have been used in the literature, both as regressors and as kernels for the cross-correlation method.

High-pass filtering. High-pass filtering has been used to reduce the effects of low-frequency noise, which can adversely affect the analysis of time courses. For example, low-frequency noise introduces correlations in the sampled magnetic resonance (MR) signal that affect the validity of statistical techniques that assume independent sampling. While filtering can be beneficial, its effects depend on the nature of the experimental paradigm. Rapidly presented event-related designs have very different frequency-domain characteristics than either block designs or widely spaced event-related designs. In this paper, we analyze the effects of high-pass filtering in different experimental paradigms by presenting models for the noise and signal components of the data.

METHODS

Experimental Methods

As described in Ollinger *et al.* (2001), data were acquired from four subjects. The same stimuli and image acquisition protocol were used here. The high-contrast flickering checkerboard was presented in a rapid event-related design with the ITIs uniformly distributed across intervals of 2.4, 4.7, and 7.1 s.

Simulation Methods

Monte Carlo simulations were used to determine the optimal values of the minimum and maximum ITIs and the optimum fraction of partial trials. We define optimum values to be those that minimize the mean variance of points in the time courses, that minimize the root-mean-square (RMS) variation of the variance, and that minimize the RMS value of the correlation coefficients. These metrics were computed for each simulated data set and then averaged across data sets. An experiment with three stimuli combined in two types of trials was simulated. Each trial consisted of the first stimulus (which we call the cue) followed after a two-frame interval by either the second or third stimulus (which we call the targets). Randomly generated design matrices for measured data consisting of 2048 points of data were created for paradigms with a range of partial trial fractions and ITIs. The ITI is defined as the interval between the end of one trial and the beginning of the next. In the first simulation run, the

minimum ITI (ITI_{min}) was varied from zero to five TRs while the maximum ITI was varied from ITI_{min} + 1 to 7 TRs with a partial-trial fraction of 0.25. Note that the data are discretely sampled and that we are considering data measured under the null hypothesis. Since the sampling rate only affects the representation of the hemodynamic response, which is absent under the null hypothesis, the relevant unit of time is a single sample, i.e., a single TR. Therefore, the statistical results do not depend on the duration of the TR, although calculations of the magnitude of the response obviously do.

The first simulation run was repeated twice: once with a uniform distribution of ITIs and again with an exponential distribution. The exponential distribution was formed by assigning one-half of the ITIs to the minimum ITI (ITI_{min}), one-fourth to ITI_{min} + 1, one-eighth to ITI_{min} + 2 and so forth until the maximum ITI was reached. All remaining trials were assigned to the maximum ITI. Note that this method forces the exponential and uniform distributions to be the same if there are only two ITIs. The second simulation run varied the partial-trial fraction from 0.1 to 0.8 for ITI ranges of 0–3, 1–4, and 2–5 for both uniform and exponential ITI distributions. The first run was then repeated for the partial-trial fraction that yielded the best statistical properties. All simulations were repeated for time course lengths of 8 and 16 TRs. For a TR of 2.5 s, the 8 TR time courses would sample the first 20 s of the BOLD response. The 16 TR time courses would sample the first 40 s. A pseudo-random number generator was used to generate 25 realizations of stimuli meeting each of these specifications. The design matrix A was computed for each realization.

As in all linear models with nonorthogonal regressors, the covariance matrix is strongly affected by the experimental design. In the case of rapidly presented FMRI studies, this source of covariance is dominant (Ollinger and McAvoy, 2000). Since the general linear model assumes that the noise in the data is white and Gaussian, the covariance matrix of the estimates $\hat{b}$ is proportional to $\boldsymbol{K}_{all} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}$ (Beck and Arnold, 1977). The covariance matrix of the effects of interest is given by a submatrix $\boldsymbol{K}$ of $\boldsymbol{K}_{all}$. The variances of the estimates are given by the diagonal terms k_{ii} and the correlation coefficients are given by the normalized off-diagonal terms, i.e., by $k_{ij}/\sqrt{k_{ii}k_{jj}}$. The metrics described above were computed from these variances and correlation coefficients.

Statistical Analysis

Statistical maps were generated by cross-correlating the time courses with models for the hemodynamic response and then forming t statistics. Several models for the hemodynamic response have been proposed. One is a gamma function (Boynton *et al.*, 1996), which can be modified by adding a variable delay (Dale and Buckner, 1997). An alternative model uses a Gaussian function (Clark *et al.*, 1997; Zarahn *et al.*, 1997). A more complex model uses a set of three gamma functions and their derivatives to account for undershoot at the tail of the function and to desensitize the model to slight changes in temporal delay (Friston *et al.*, 1998). This model is perhaps the most accurate, but is of limited usefulness in random effects models because it yields six magnitudes rather than one, and these cannot be combined to form a single, normally distributed value that represents the magnitude of the activation. A simplified version, referred to in the SPM package (Wellcome Department of Cognitive Neurology, London: http://www.fil.ion.ucl.ac.uk/spm) as the "canonical" model and referred to here as the "SPM canonical" model, is the difference of two gamma functions: one to model the peak of the response and the other to model the postpeak undershoot.

The sensitivity of the statistical maps to the model for the hemodynamic response was characterized by computing statistical maps using the two most widely used methods: The SPM canonical model and the delayed gamma function model (Dale and Buckner, 1997). Both were convolved with a boxcar function to model the duration of the stimulus. Details are given in the Appendix. Each model was used two ways: as a regressor in the design matrix and as a kernel in the cross-correlation of the model function with the time courses estimated as described in Ollinger *et al.* (2001). All models included a high-pass filter with a cutoff frequency of 0.014 Hz (i.e., the high-pass filter removes the lowest four frequencies in the data). This filter was implemented by including sine and cosine functions at each frequency below the cutoff frequency as regressors in the design matrix. The variance was estimated from the residuals (Friston *et al.*, 1995a, Friston *et al.*, 1995b; Worsley and Friston, 1995), and t statistics were formed as the ratio of the magnitude to its standard deviation. F statistics were computed using the extra sum-of-squares approach (Beck and Arnold, 1977). These t and F statistics were then "Gaussianized," i.e., transformed to normally distributed statistics with the same significance probabilities. The unsmoothed statistical maps were corrected for multiple comparisons using a Bonferroni correction.

Noise and Signal Power Spectrum

The noise power spectrum was determined by scanning a subject under two conditions. In the first, 1280 volumes were collected in a fixation state with a TR of 2.5 s and a voxel size of 3 × 3 × 3 mm. In the second, the subject viewed a flickering checkerboard in a block

paradigm. Statistical maps were computed as described above for the activation data. A Bonferroni threshold was then applied to yield a map of activated voxels in visual cortex. The power spectra were then computed at each voxel for each run of the fixation data. These were then averaged across runs to yield average power spectra at each voxel and then averaged again across activated voxels to yield a low-noise estimate of the noise power spectra. This power spectrum is representative of regions that are predominantly gray matter in visual cortex at rest. We assume that this is representative of the power spectrum of the noise during a task-induced activation.

RESULTS

Correlations in the Estimated Time Courses

Correlations among points in the estimated time courses arise from two sources: correlations in the measured data and correlations induced by the design. We consider correlations induced by the experimental design here. Note that these correlations quantify the relationship between the *noise* at a point in one estimated time course and the *noise* at points in the same or other estimated time courses. They do not describe correlations among the signals. Under the assumptions of the linear model, the estimated time courses are always equal to the true time course corrupted by this correlated noise.

The structure of these correlations can be seen in the profile through a covariance matrix for the simulated experiment shown at the top of Fig. 1. Each trial of the compound trial experiment (open circles) involved a cue (effect 1) that was followed after two TRs by one of two targets (effect 2 or 3). Each trial of the rapid event-related experiment contained a single stimulus that could be of three different types (effect 1, 2, or 3). This profile is taken through row 8, which is the eighth point ($t = 8$) of the first effect in the compound trials. The value at column eight is equal to 1 since any point is perfectly correlated with itself. There are two prominent features. First, in the compound trial experiment, there are negative peaks at rows corresponding to $t = 6$ of each target effect and positive peaks at $t = 8$ of each target effect. The negative peaks are explained by noting that the eighth point of the cue overlays the sixth point of each target response, since the target effect begins two TRs after the cue effect. Any noise that is attributed to the eighth point of the cue must be taken away from these points in the target response, which results in a negative correlation coefficient.

The positive peaks in both the compound trial and event-related experiments in Fig. 1 are more difficult to explain. They can be understood by considering the simpler case of a rapidly presented event-related design with three effects. If the stimuli are widely spaced and presented in such a way that each point in the data contributes to one and only one effect, the columns corresponding to the estimated effects will be collinear with the columns used to estimate the mean. Such a design does not yield a unique solution, but it can be constrained by forcing the regressors for the effects of interest to sum to 0 at each row of the design matrix. This yields unique time courses but with an added constant. This constant is of no concern if a contrast (which must sum to 0) is used to generate statistical maps because the constant term will cancel. However, this constant represents a significant, constant correlation among time points. As the trials become more closely spaced, this correlation is no longer constant, as is shown in the bottom panel of Fig. 1. Instead, the correlation coefficients become smoothly decaying functions with peaks at the same point in each effect (i.e., the cross-correlations with frame 3 for effect 1 are

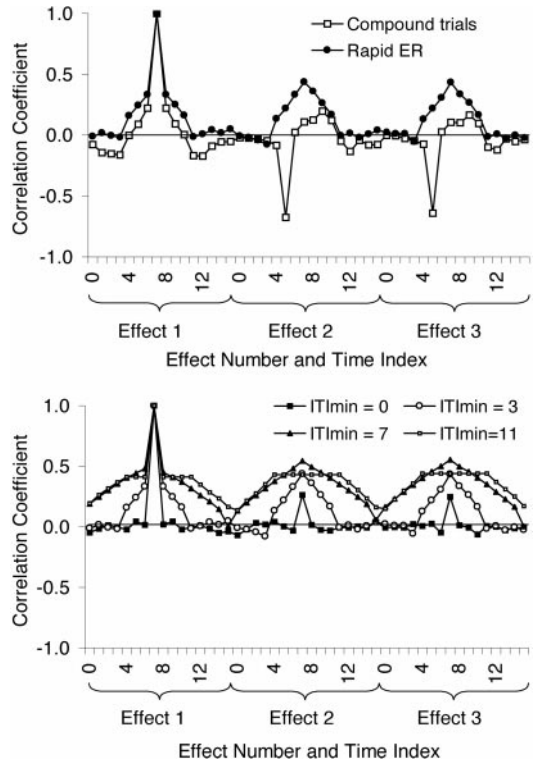

FIG. 1. Profiles through the covariance matrix at the row corresponding to the eighth point of the first effect. Each plot shows the correlation coefficient of this point of the estimated time course with each point of every other effect. The graph at the top shows the correlations for the compound event-related study (open squares) and a rapidly presented event-related study. The minimum ITI in each case was three TRs. The graph at the bottom shows the correlations for a rapidly presented paradigm for four values of the minimum ITI. This demonstrates how reducing the minimum ITI decorrelates the estimates.

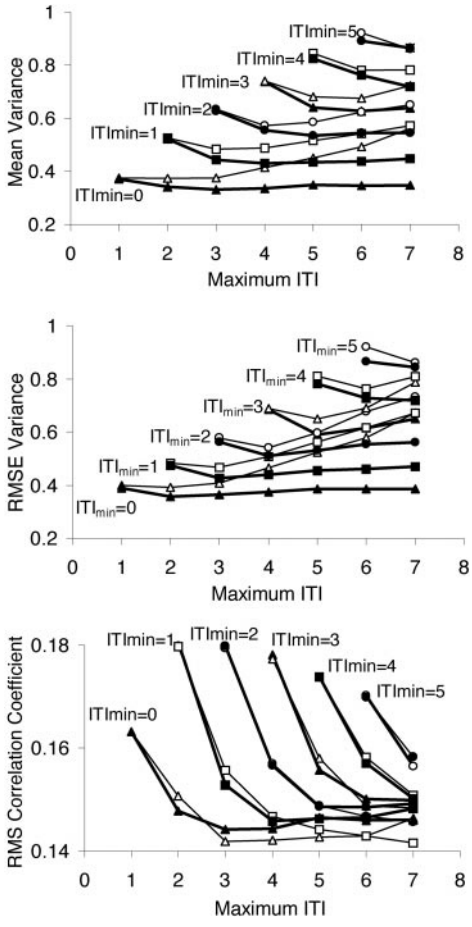

FIG. 2. Plots of mean variance, RMS error of the variance, and RMS correlation coefficient vs maximum ITI. The minimum ITI was varied from zero to five TRs while the maximum ITI was varied from ITI_{min} + 1 to 7 TRs. The light lines with open symbols represent studies with uniformly distributed ITIs and the heavy lines with closed symbols represent studies with exponentially distributed ITIs.

at a maximum for frame 3 of effects 2 and 3). These correlations will not cancel when a contrast is applied and may alter the shape of the time course in ways that are difficult to predict. Moreover, the value of the noise at adjacent points should change smoothly. Referring to the plots in Fig. 5 of Ollinger *et al.* (2001), we see that there is very little spike-like noise in the tails of the estimated BOLD responses. For any given subject, the tails look smooth and interpretable despite the fact that the variance is appreciable. Therefore, noise could be interpreted as a smooth change in the shape of the response.

One might think that removing the mean from the data prior to the estimation step rather than as part of it would eliminate some of this correlation. We compared precorrecting the data with correcting as part of the estimation procedure by comparing time courses for the high-contrast study. The maximum difference in the percentage change was 0.0264 for a time course with a peak value of 1.36. This difference is negligible.

Simulation Results

Simulation results for mean variance, RMS variation of the variance, and RMS correlation coefficient are shown in Fig. 2. The results were the same for partial-trial fractions of 0.25 and 0.4. Several conclusions can be drawn. First, exponentially distributed ITIs yield lower variance. This is not surprising, since the exponential distribution concentrates trials at shorter ITIs, which yields more trials per study. Second, the mean and RMS variation of the variance are relatively insensitive to the maximum ITI for exponentially distributed ITIs. Again, this is not surprising since most trials with the exponential distribution are at very short ITIs regardless of the maximum ITI. The variance increases as the maximum ITI increases for uniform ITIs because more trials have a long ITI, thereby decreasing the total number of trials. The RMS correlation coefficients are roughly equivalent for uniform and exponential distributions. In both cases, they are near their minimum values if there are at least three distinct ITIs.

A plot of these same statistics against the fraction of partial trials is shown in Fig. 3. This plot used ITIs exponentially distributed over the range of one to four TRs. The curves all had the same shape for all ITIs tested (zero to three, one to four, and two to five, both uniformly and exponentially distributed). It can be concluded that if variance were the only consideration, the optimum fraction of partial trials would be 0.4. Behavioral considerations usually require a lower fraction of partial trials and Fig. 3 indicates that the variance functions begin to bottom out for fractions of 0.2–0.3. Previous studies have found reliable separation of the components of a compound trial using fractions of 0.25 (Shulman *et al.,* 1999) and 0.20 (Corbetta *et al.,* 2000).

These simulation results were computed from estimates of 8-point time courses. Repeating the simula-

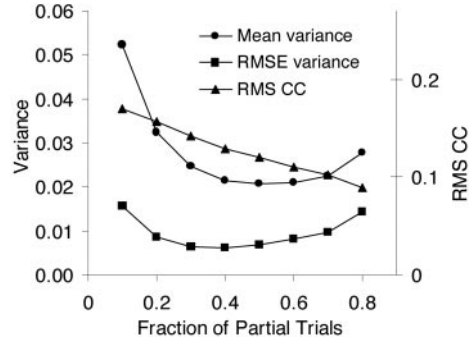

FIG. 3. Mean variance, RMS variation of the variance, and RMS correlation coefficient vs fraction of partial trials for an exponential distribution of ITIs of zero to three TRs.

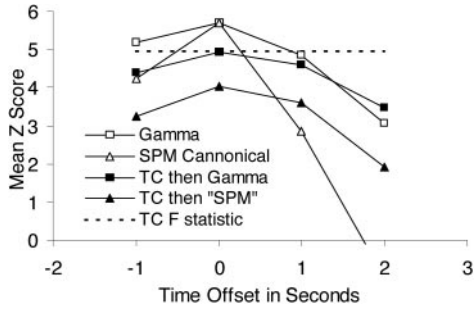

FIG. 4. Mean estimated *Z* scores vs time shift for five cases: Modeling the hemodynamic response in the design matrix with a gamma function (Gamma); modeling the hemodynamic response in the design matrix with the SPM canonical model (SPM Canonical); estimating the response time course and then cross-correlating with a gamma function (TC then Gamma); estimating the response time course and then cross-correlating with the SPM canonical model (TC then "SPM"); and finally, computing the time courses and their Gaussianized *F* statistic (TC *F* statistic).

tions for 16-point time courses yields curves with the same shapes. However, the mean variance increases by 13%, the RMS variation of the variance decreases by 11%, and the RMS correlation coefficient decreases by 25%. Therefore, estimating 16 points would decrease *z* scores by an average of 5.9%, but would yield estimates with more homogeneous variances.

Statistical Analysis

The sensitivity of the statistical maps to the model for the hemodynamic response was quantified by *z* statistics averaged over V1 for the high-contrast study. The SPM canonical model and the delayed gamma function model were evaluated at four different delays: the nominal value of 2 s and shifts from this value of −1, 1, and 2 s. The means across subjects are shown in Fig. 4. If the delay is known exactly, the SPM canonical model yields the highest *z* statistics. The performance of the canonical model falls off rapidly, however, as the delay of the true BOLD response deviates from that of the model response. Both models yield lower *z* statistics when they are formed by cross-correlating the model response with the estimated time course of the BOLD response rather than by using the model as a regressor in the design matrix. Note, however, that this disadvantage decreases and may actually reverse if the delay is not optimal. With the cross-correlation technique, the gamma function model yields higher *z* statistics and is more robust to time shifts than the canonical model. The *F* statistic approach, which does not make any shape assumptions, yields *z* scores equivalent to the highest obtained with either cross-correlation method.

Models for the Signal and Noise

Signal Model

A frequency domain representation for the signal is easy to develop for paradigms with regularly spaced stimuli, such as block designs and widely spaced event-related studies. Any periodic function can be represented by the sum of sinusoids at frequencies that are a multiple of the paradigm frequency. For example, a block design with task/control blocks of 40 s each would be represented by the weighted sum of sinusoids at frequencies of 0.0125, 0.0250, 0.0375 Hz, etc., where the weights decrease rapidly with frequency. Similarly, a widely spaced event-related study with stimuli presented every eight TRs (20 s) would be represented by sinusoids at frequencies of 0.05, 0.10, 0.15 Hz, etc. A representation for the paradigms proposed here can be inferred by considering a rapidly presented event-related study and then assuming that the power spectrum for a compound event-related study will be similar. We model the stimulus as a train of impulses, $s(t)$, and assume that brain hemodynamics are well modeled by a linear system with an impulse response, $h(t)$, the hemodynamic response function. The BOLD response is therefore given by $s(t) * h(t)$, where "*" denotes convolution. Moreover, if the ITIs follow an exponential distribution, i.e., half the ITIs are equal to 0, one-fourth of the ITIs are equal to 1 TR, one-eighth of the ITIs are equal to 2 TRs, etc., then the stimulus $s(t)$ is well modeled as a homogeneous Poisson process (Snyder and Miller, 1991). By Campbell's theorem (Snyder and Miller, 1991), the BOLD response has a power density spectrum given by $P(\omega) = \lambda|H(\omega)|^2$, where λ is the intensity of the Poisson process, $H(\omega)$ is the Fourier transform of $h(t)$, and $|\cdot|$ denotes magnitude. If $h(t)$ is modeled by the gamma function $h(t) = At\exp(-\alpha t)$ (Boynton *et al.*, 1996), the power spectrum

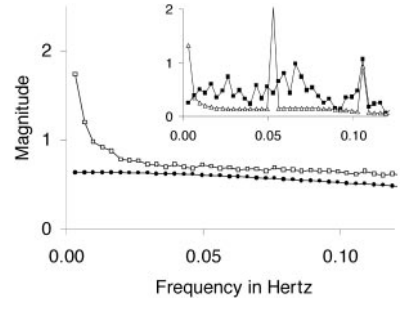

FIG. 5. Magnitude of the measured noise power (open squares) and the computed signal power (closed circles) vs temporal frequency. The noise power was measured during a fixation task. The signal power was computed analytically for a rapid event-related design with an exponential distribution of ITIs and minimum ITI of one TR. The signal power for a widely spaced design (open triangles) and the rapidly presented design used in the experimental study presented here (closed squares) are shown in the inset.

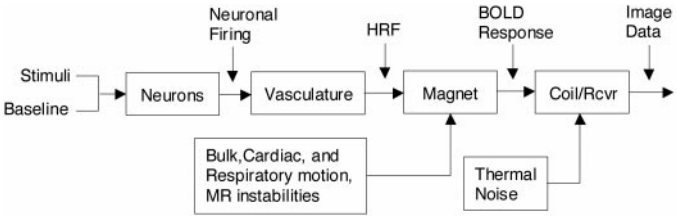

FIG. 6. Noise sources in a typical FMRI study.

is given by $P(f) = \lambda|1/(\omega^2 + \alpha^2)|^2$, where $\omega = 2\pi f$. A typical value of α is 0.8 s (Dale and Buckner, 1997). The resulting power spectrum in Fig. 5 shows that there is significant signal power at all frequencies, including very low ones. For example, the summed responses to overlapped trials do not return to zero and will therefore have a nonzero mean, so there is even signal power at a frequency of zero.

There are two shortcomings to this analysis. First, it does not include the postpeak undershoot of the hemodynamic response. This causes an underestimate of the signal power at very low frequencies. Second, it assumes that the hemodynamic response has infinite support. This leads to a slight overestimate of the signal power at very low frequencies. Nevertheless, the general shape is representative of the power spectrum of the measured BOLD response.

Noise Model

The evolution of the noise from the neuronal level to the received signal is modeled by a linear system as shown in Fig. 6. The inputs to the system at the neuronal level are the presented stimuli and low-frequency fluctuations in neuronal firing that may arise from functional connectivity (Biswal *et al.,* 1995). Coupling between neuronal firing and hemodynamics can be modeled as the output of a linear system whose kernel is the hemodynamic response function. This implies that a short burst of neuronal firing causes the transient increase in oxygenated blood described by the hemodynamic response. In the presence of the magnetic field, this HRF induces the BOLD signal. At this stage, noise due to bulk subject motion, cardiac pulsatility, respiratory motion, and various other noise sources in the MR signal is introduced. The BOLD response is observed through the coil and receiver, which add thermal noise to the signal.

Two of these noise sources are easy to deal with. First, the noise due to baseline fluctuations in neuronal firing is quite small since it is of the same order of magnitude as the firing changes due to functional connectivity, and these are small (Biswal *et al.,* 1995; Xiong *et al.,* 1999) and can be neglected. Second, thermal noise added to the RF signal in the coil and receiver is uncorrelated, stationary, and Gaussian to a very good approximation in regions of the image with a mean much greater than zero and therefore satisfies the assumptions of the linear model. Noise introduced into the BOLD signal (as distinguished from the two previous noise sources) is more difficult to characterize. Noise due to respiratory motion occurs at a frequency of approximately 0.1 Hz. It is not usually modeled. Noise due to cardiac pulsatility is usually sampled at rates too low to adequately represent it. For example, if the pulse rate is 60 bpm, the Nyquist sampling interval will be 0.5 s. A typical TR of 2.5 s undersamples this signal by a factor of five. The effect will be to alias the cardiac signal.

An empirically determined noise power spectrum (see Methods) is shown in Fig. 5. As has been observed by others (Zarahn *et al.,* 1997), the spectrum can be modeled as the sum of a $1/f$ decay and a constant. The constant term is assumed to correspond to the thermal noise, and the low-frequency noise is caused by the other effects mentioned above (Zarahn *et al.,* 1997). Since uncorrelated, normally distributed data yield a uniform power spectrum, the deviation of the spectrum shown in Fig. 5 from a uniform distribution can be interpreted as correlations among points in the measured data. This deviation stems from the physiological and MR noise which is concentrated at low frequencies (Zarahn *et al.,* 1997).

Several approaches have been suggested for dealing with correlated noise. Optimal filtering can be used to prewhiten the data (Buonocore and Maddock, 1997), but this requires additional data collection in order to define the filter. Another approach appeals to the matched-filter theorem to justify smoothing with the HRF (Friston *et al.,* 1995a,b). This, however, only deals with the baseline noise. A third approach is to model the correlations as a first-order autoregressive process (Burock, 1998; Purdon and Weisskoff, 1998). The most widespread approach is to high-pass filter the data. The goal of filtering is to remove noise at frequencies that do not contain the signal. Since the variance is computed from the residuals of the linear model, removing these frequencies decreases the variance, thereby increasing the t statistics estimated from the model. Filtering does not reduce the incidence of false positives because it does not remove noise at the paradigm frequency. False positives can be reduced by designing experiments so that the signal energy occurs predominantly at frequencies that contain relatively small amounts of noise.

Effects of High-Pass Filtering

As described above, block paradigms often concentrate most of their energy at frequencies on the order of 0.0125 Hz. Figure 5 shows that this frequency lies near the $1/f$ portion of the noise spectrum. This explains the

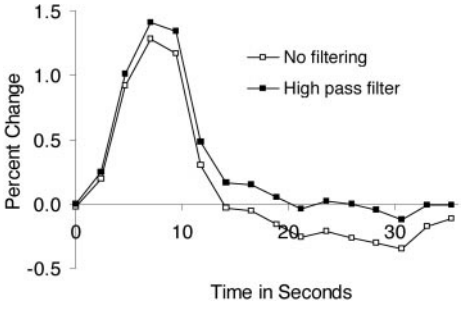

FIG. 7. Mean time course of the high-contrast response estimated from the high-contrast experiment using two different models. The first incorporated a high-pass filter that modeled the lowest four frequency components while the other did not.

susceptibility of block designs to physiological noise (Zarahn *et al.,* 1997). As shown in the inset to Fig. 5, widely spaced event-related designs would typically have the most energy at roughly 0.055 Hz. This frequency lies outside the low-frequency noise peak, which explains the reduced susceptibility of these studies to physiological noise. In both cases, one would expect high-pass filtering to improve statistical power by reducing the variance since it should decrease the estimated variance while leaving the BOLD signal unaffected.

The situation is more complex for rapidly presented event-related designs. These designs spread the signal energy across the entire spectrum as shown in Fig. 5 for the analytical model. This has two effects. First, it deemphasizes low-frequency noise so the estimation is relatively less sensitive to these frequencies; but second, it places some signal energy at the noisiest frequencies. Therefore, high-pass filtering will remove part of the BOLD signal. The power spectrum for the experimental design used here is shown in the inset to Fig. 5. It is neither as smooth nor as flat as that of the analytical model. This results from the finite length of the experiment and from the nonexponential distribution of the ITIs.

The effect on the estimates can be understood by considering the postpeak undershoot. This undershoot has a duration of roughly 90 s (Fransson *et al.,* 1998; Mandeville *et al.,* 1998) and is therefore represented predominantly by frequency components at approximately 0.01 Hz. One would expect it to be strongly affected by the high-pass filter. This hypothesis was tested by analyzing the data both with and without a high-pass filter with a cutoff at 0.014 Hz. As shown in Fig. 7, the high-pass filter eliminates the undershoot from the estimated time course. This implies that models for the hemodynamic response function should not include the undershoot if the data are high-pass filtered.

The merit of the high-pass filter was investigated by analyzing how filtering affected *z* scores for the high-contrast studies for five methods of computing statistical maps. For the first two methods, *Z* scores were obtained by coding the SPM canonical and gamma function regressors into the design matrix as regressors. The other three methods estimated time courses and their associated *F* statistics for these conditions: (1) rapidly presented with 8 estimated points, (2) rapidly presented with 16 estimated points, and (3) widely spaced with 8 estimated points. The differences between *z* statistics computed with and without high-pass filtering are shown in Fig. 8. Three of the methods do not model the undershoot as part of the hemodynamic response: the gamma function regressor and the two methods that estimate 8-point time courses (these truncate the undershoot). In each of these cases, the *z* scores increased for each subject when the high-pass filter was included. The SPM canonical model accounts for the undershoot twice: as part of the hemodynamic response model and in the high-pass filter. Again, there were consistent improvements with the high-pass filter for three of four subjects, but only a very small improvement for the fourth. We speculate that in this subject, the undershoot matched the undershoot in the model well enough that the high-pass filter made little difference. The case of the 16-point time course was more interesting. For one subject, high-pass filtering *reduced* the *z* statistic, for another there was a small improvement, and for the other two it increased the *z* statistic. Interestingly, the two subjects for whom the *z* statistic increased (subjects 2 and 4) were also the subjects that appeared to have more motion artifact. We infer from this that high-pass filtering the data can degrade sensitivity in cooperative subjects under relatively noise-free conditions, but can improve performance under noisy conditions. The changes in *z* statistics were small, typically on the order of 0.25, and the mean *z* score was between 5 and 5.5 depending on the

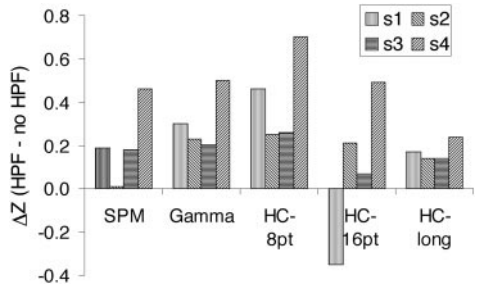

FIG. 8. The difference between statistical maps computed with and without a high-pass filter for four subjects, s1 through s4, and five processing methods. The "SPM" and "Gamma" used the SPM canonical model and a gamma function, respectively, as regressors. The other three approaches all computed time courses and their associated Gaussianized *F* statistics. The first two estimated 8- and 16-point time courses from the rapidly presented high-contrast study, while the third estimated an 8-point time course from the widely spaced high-contrast study.

computational method. This amounts to a 5% change in the estimated magnitude, which is unlikely to make an appreciable difference in most studies.

We draw the following conclusions about high-pass filtering. For block designs and widely spaced event-related designs, it removes correlations from the data and reduces variance estimates at the cost of a reduced number of degrees of freedom (Friston *et al.*, 1995a,b; Worsley and Friston, 1995). Since most FMRI experiments have hundreds to thousands of degrees of freedom, this reduction does not have a significant effect on sensitivity. Filtering may also be useful for rapidly presented event-related designs. It again removes correlations from the data and reduces variance estimates, although it also removes some signal. In particular, since filtering removes the postpeak undershoot, it is best done in conjunction with a model for the hemodynamic response that does not include that undershoot.

DISCUSSION

Correlations in the Estimated Time Courses

One might think that time courses estimated with FMRI could be analyzed in the same way as those acquired with single-unit recordings. This, however, is not the case. In single-unit recordings, the noise is uncorrelated across time points and has approximately equal variances at each time point. This gives the noise component of the data a spike-like appearance, which can be easily recognized. In FMRI, however, the noise across time points in the estimated time courses shows a high degree of correlation. This follows from the simulation results as well as the empirical evidence in the time courses. These correlations imply that the noise in estimated time courses will be smooth rather than impulsive and could be interpreted as features of the BOLD response. As one would expect, this smooth noise is greatly reduced in across-subject averages. This suggests that estimated time courses should be analyzed with caution in single subjects. Monte Carlo results show that correlations between estimated time points of the same effect are minimized by using short minimum ITIs.

Optimal Experiment Design

Optimal designs are those that yield the largest estimated magnitudes with the best statistical properties while satisfying the behavioral constraints of the experiment. We define the best statistical properties as low variance, equal variance across effects, and minimum correlation among effects. The simulations show that the first two properties are best achieved with short ITIs that follow an exponential distribution, probably because these designs maximize the number of trials. The overall magnitude of correlation is minimized by using at least three distinct ITIs. Although the correlation among effects is relatively independent of the choice of ITIs, shorter minimum ITIs reduce correlations among points within a given time course, as noted above. Short minimum ITIs yield good statistical properties, but they may also produce nonlinear interactions among effects. In particular, the presence of modest nonlinearities (Miezin *et al.*, 2000) suggests that exponential distributions which produce a succession of trials involving short ITIs may be undesirable. Since response magnitude increases for longer minimum ITIs (Miezin *et al.*, 2000), the optimum range of ITIs is likely to be a minimum of one to two TRs (2.4–4.7 s) and a maximum of three to four TRs (7.1–9.4 s).

Low variance, equal variance across effects, and minimum correlation among effects are all minimized by using 40% partial trials. Behavioral constraints, however, will generally require a lower fraction (Shulman *et al.*, 1999; Corbetta *et al.*, 2000). A fraction of 0.25 yielded reliable separation of the low-contrast and high-contrast responses in the study presented here, and good results have also been obtained in previous studies with fractions of 0.25 (Shulman *et al.*, 1999) and 0.2 (Corbetta *et al.*, 2000).

Statistical Analysis

When the model for the hemodynamic response was accurate, coding the model function into the design matrix as a regressor yielded a higher *z* score than estimating a time course and cross-correlating that time course with the model function. This advantage decreased, however, and sometimes reversed, when the delay time was inaccurate. Both model functions yielded results that were affected by errors in the delay time. This has the undesirable effect of confounding the delay of the response with its measured magnitude. The delayed gamma function model was more robust than the SPM canonical model in this respect.

The relevance of this sensitivity to the delay time can be assessed by considering the standard deviation of the time-to-peak of the hemodynamic response, which is approximately 1.1 s (Aguirre *et al.*, 1998) for a 500-ms visual stimulus. This implies that the range of values tested here is likely to occur in many experiments. Moreover, the stimuli presented here had a much longer duration of 4.7 s, which leads to much wider hemodynamic responses. Magnitudes computed from these wider hemodynamic responses would be expected to be more robust with respect to variations in timing than the relatively narrow hemodynamic responses in Aguirre *et al.* (1998). Therefore, we would expect experiments using brief stimuli to be even more

sensitive to changes in the delay parameter than the data presented here suggest.

These results suggest that the model functions should only be used as regressors in the design matrix if they are known to accurately model the form of the BOLD responses in all regions of interest. In the absence of this detailed prior information, response magnitude will be confounded with the timing and width of the response. Estimating the time courses reduces the possibility of missing significant activations that depart from the standard shape or delay and prevents biased analyses that favor conditions that yield hemodynamic responses that match the model function.

Results for the hemodynamic response in V1 to a passive visual stimulus may not be strictly representative of the response of other areas or the response during demanding cognitive tasks. This seems particularly likely if the task has an extended duration and engages cognitive processes of different durations at different time points during the task. It might therefore make more sense to estimate the time course at each voxel and then conduct a statistical analysis on the time course functions. Otherwise, the analysis is biased toward those areas and conditions that yield canonical time courses. The use of strong shape assumptions may only be advisable for areas and tasks, which have already been well characterized (i.e., V1 responses during passive sensory stimulation). Although the *F* statistic, which does not make any assumption about shape, yielded lower *z* statistics than those obtained by coding the model functions into the design matrix as regressors, their equivalent *z* scores were equal to the highest values yielded by the correlation of time courses with model functions. Moreover, when the delay of the hemodynamic response was off by as little as 1 s, the *F* statistics were equivalent to the highest obtained using model functions as regressors to obtain *t* statistics.

This analysis holds for data analyzed at the level of a single subject. At this level, any modeling choice can affect either the magnitude or the variance of an effect. Most experiments, however, are analyzed across subjects. In these analyses, magnitudes of activation are extracted from each subject and then compared in second-level analyses using paired *t* tests or analysis of variance (ANOVA) tests. These second-level analyses are based on the across-subject variance, which may not be dominated by the variance at the single-subject level. Therefore, any aspect of the model that affects the magnitude of the response should affect the group and individual analyses equally while any aspect that only affects the within-subject variance term may or may not have a strong impact on the group analyses. Since the choice of model functions primarily affects the magnitude of the estimated response, we expect these results to carry over to group analyses.

The need for a model for the hemodynamic response can be eliminated if the time courses are analyzed with ANOVA methods (Shulman *et al.*, 1999; Corbetta *et al.*, 2000). This is done by treating time as a main effect with levels corresponding to each estimated point of the BOLD responses. If the experiment has one task with two levels, the resulting ANOVA would have two main effects (task and time) and one interaction term. The main effect of time would test the hypothesis that estimated time courses averaged across tasks do not significantly vary from zero. The interaction term would test whether the time course varied significantly across the two tasks.

A fundamental assumption of ANOVAs is that the variance is homogeneous, i.e., that the variances are equal and that data points are uncorrelated. The time courses discussed here violate this assumption. Therefore, ANOVAs can only be used if the degrees of freedom of the ANOVA are adjusted appropriately (Box, 1954; Ollinger and McAvoy, 2000). This incurs a loss of statistical power that becomes more severe as the variance becomes less homogeneous. Therefore, proper choice of the timing parameters of a study can have a significant effect on its statistical power.

High-Pass Filtering

The power spectrum of the data changes significantly with the type of design and the timing of the stimuli. Block designs concentrate the signal power at low frequencies; widely spaced event-related designs concentrate it at much higher frequencies; and rapidly presented event-related designs spread it throughout the power spectrum. Since the noise power spectrum peaks at low frequencies (Fig. 5), the event-related designs are less sensitive to artifacts due to low-frequency noise than block designs. Therefore, these designs may be preferable for studies where low-frequency noise due to patient motion is expected to be high, such as studies in patient populations and studies that require speaking in the scanner.

High-pass filtering is beneficial for block designs and widely spaced event-related designs, since it removes $1/f$ noise from the data, thereby reducing the variance estimates and whitening the data. For similar reasons, filtering can also be useful for rapidly presented event-related designs. In addition, however, the signal in rapid event-related designs has power at low frequencies, so filtering affects the estimated time course. As shown in Fig. 7, the contribution to the signal due to the postpeak undershoot is largely removed. This leads to better fits by models that do not accurately account for the undershoot.

APPENDIX

Two models for the hemodynamic response to a brief stimulus were used: a single gamma function (Boynton *et al.*, 1996) and the difference of two gamma functions (Friston *et al.*, 1999). These models were modified to account for the relatively long duration of the stimuli used in this paper by convolving them analytically with a boxcar function. Details are given below.

Gamma Function Model

The hemodynamic response is modeled as the convolution $h(t) = g(t) * b(t)$, where $g(t)$ is the delayed gamma function given by

$$g(t) = \alpha(t - \tau)e^{-a(t-\tau)} \tag{1}$$

The variable α, which determines the width of the hemodynamic response, is fixed at 1.25 s (Dale and Buckner, 1997). The variable τ, which represents the delay of the response, defaults to a value of 2 s. The second function in the convolution, $b(t)$, is the boxcar function given by

$$r(t) = \begin{cases} 1/T, & 0 \le t \le T \\ 0, & \text{elsewhere,} \end{cases} \tag{2}$$

where the variable T is the assumed duration of neuronal firing. For studies where the stimulus is sustained, such as the flickering checkerboard used in the studies presented here, T is the duration of the stimulus presentation. In most studies, T is the duration of neuronal firing and is on the order of tens or hundreds of milliseconds. In those cases, we assume that T is small compared with the TR and model $h(t)$ by an impulse function. When T is not small, the convolution is given by

$$h(t) = \begin{cases} 0, & t < 0 \\ e^{-\alpha t}(\alpha^2 t^2 + 2\alpha t + 2), & 0 \le t \le T \\ e^{-\alpha(t-T)}(\alpha^2(T-t)^2 + 2\alpha(T-t) + 2) & \\ \quad - e^{-\alpha t}(\alpha^2 t^2 + 2\alpha t + 2), & T < t \end{cases} \tag{3}$$

SPM Canonical Model

The canonical model used in the SPM package (Friston *et al.*, 1999) is defined as

$$g(t) = \gamma_6(t) - 1/6\gamma_{16}(t) \tag{4}$$

where $\gamma_n(t)$ is the gamma function defined by

$$\gamma_n(t) = \frac{1}{\Gamma(n)} t^{n-1}e^{-t}. \tag{5}$$

Convolving with a boxcar function as before yields the result that

$$h(t) = \begin{cases} 0, & t < 0 \\ \displaystyle\int_0^t \gamma_6(t-u)du - \frac{1}{6}\int_0^t \gamma_{16}(t-u)du, & 0 \le t \le T \\ \displaystyle\int_0^T \gamma_6(t-u)du - \frac{1}{6}\int_0^T \gamma_{16}(t-u)du, & T \le t \end{cases}, \tag{6}$$

where the integral is given by

$$\int_0^t \gamma_n(t-u)du = \frac{e^{-t}}{\Gamma(h)}\left(\sum_{i=0}^{n}\frac{n!t^{n-i}}{(n-i)!}\right)\Bigg|_t^{t-T}. \tag{7}$$

ACKNOWLEDGMENTS

This work was supported by Grants EY00379 and EY12148 from the National Institutes of Health and by a grant from the McDonnell Center for Higher Brain Function. The authors thank Francis Miezin for the data used to compute the noise power spectrum, Tom Conturo and Erbil Akbudak for the MRI pulse sequence development, and Avi Snyder for the development of the preprocessing software. We also thank Randy Buckner and Mark McAvoy for a careful reading of the manuscript. Finally, we thank the referees for their useful suggestions.

REFERENCES

Aguirre, G. K., Zarahn, E., and D'Esposito, M. 1998. The variability of human, BOLD hemodynamic responses. *NeuroImage* **8:** 360–369.

Beck, J. V., and Arnold, K. J. 1977. *Parameter Estimation in Engineering and Science.* Wiley, New York.

Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* **34**(4): 537–541.

Box, G. E. P. 1954. Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and correlation between errors in the two-way classification. *Ann. Math. Stat.* **25:** 290–302.

Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* **16**(13): 4207–4221.

Buckner, R. L., Bandettini, P. A., O'Craven, K. M., Savoy, R. L., *et al.* 1996. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging [see comments]. *Proc. Natl. Acad. Sci. USA* **93**(25): 14878–14883.

Buckner, R. L., Goodman, J., Burock, M., Rotte, M., *et al.* 1998. Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. *Neuron* **20**(2): 285–296.

Buonocore, M. H., and Maddock, R. J. 1997. Noise suppression digital filter for functional magnetic resonance imaging based on image reference data. *Magn. Reson. Med.* **38**(3): 456–469.

Burock, M. A. 1998. Design and statistical analysis of FMRI experiments to assess human brain hemodynamic responses. Department of Electrical Engineering and Computer Science, p. 77. MIT, Boston.

Clark, V. P., Maisog, J. M., and Haxby, J. V. 1997. FMRI studies of face memory using random stimulus sequences. *NeuroImage* **5:** S50.

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., *et al.* 2000. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.* **3**(3): 292–297.

Dale, A. M., and Buckner, R. L. 1997. Selective averaging of rapidly presented individual trial using fMRI. *Hum. Brain Mapp.* **5:** 1–12.

Fransson, P., Kruger, G., Merboldt, K. D., and Frahm, J. 1998. Temporal characteristics of oxygenation-sensitive MRI responses to visual activation in humans. *Magn. Reson. Med.* **39**(6): 912–919.

Friston, K., Holmes, A., Poline, J., Grasby, P., *et al.* 1995a. Analysis of fMRI time series revisited. *NeuroImage* **2:** 45–53.

Friston, K., Holmes, A., Worsley, K., Poline, J., *et al.* 1995b. Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* **2:** 189–210.

Friston, K., Worsley, K., Frackowiak, R., Mazziotta, J., *et al.* 1994. Assessing the significance of focal activations using their spatial extent. *Hum. Brain Mapp.* **1:** 214–220.

Friston, K. J., Holmes, A. P., and Ashburner, J. 1999. *Statistical Parametric Mapping (SPM).* http://www.fil.ion.ucl.ac.uk/spm/.

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

Kim, S. G., Richter, W., and Ugurbil, K. 1997. Limitations of temporal resolution in functional MRI. *Magn. Reson. Med.* **37**(4): 631–636.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., *et al.* 1992. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Natl. Acad. Sci. USA* **89:** 5675–5679.

Lee, A. T., Glover, G. H., and Meyer, C. H. 1995. Discrimination of large venous vessels in time-course spiral blood-oxygen-level-dependent magnetic-resonance functional neuroimaging. *Magn. Reson. Med.* **33**(6): 745–754.

Mandeville, J. B., Marota, J. J., Kosofsky, B. E., Keltner, J. R., *et al.* 1998. Dynamic functional imaging of relative cerebral blood volume during rat forepaw stimulation. *Magn. Reson. Med.* **39**(4): 615–624.

Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E., *et al.* 2000. Characterizing the hemodynamic response: Effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *NeuroImage* **11:** 735–759.

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. USA* **87:** 9868–9872.

Ollinger, J. M., and McAvoy, M. 2000. A homogeneity correction for post-hoc ANOVAs in FMRI. *Proceedings of the 6th Annual Meeting of the Organization for Human Brain Mapping.*

Ollinger, J. M., Shulman, G. L., and Corbetta, M. 2001. Separating processes within a trial in event-related functional MRI. I. The method. *NeuroImage* **13:** 210–217.

Purdon, P. L., and Weisskoff, R. M. 1998. Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. *Hum. Brain Mapp.* **6**(4): 239–249.

Schacter, D. L., Buckner, R. L., Koutstaal, W., Dale, A. M., *et al.* 1997. Late onset of anterior prefrontal activity during true and false recognition: An event-related fMRI study. *NeuroImage* **6:** 259–269.

Shulman, G. L., Ollinger, J. M., Akbudak, E., Conturo, T. E., *et al.* 1999. Areas involved in encoding and applying directional expectations to moving objects. *J. Neurosci.* **19**(21): 9480–9496.

Snyder, D. L., and Miller, M. I. 1991. *Random Point Processes 2.* Springer-Verlag, New York.

Worsley, K., and Friston, K. 1995. Analysis of fMRI time-series revisited—Again. *NeuroImage* **2:** 173–181.

Worsley, K. J., Evans, A. C., Marrett, S., and Neelin, P. 1992. A three-dimensional statistical analysis of CBF activation studies in human brain. *J. Cerebr. Blood Flow Metab.* **12:** 900–918.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., *et al.* 1995. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Mapp.* **4:** 58–73.

Xiong, J., Parsons, L. M., Gao, J. H., and Fox, P. T. 1999. Interregional connectivity to primary motor cortex revealed using MRI resting state images. *Hum. Brain Mapp.* **8**(2–3): 151–156.

Zarahn, E., Aguirre, G., and D'Esposito, M. 1997. A trial-based experimental design for fMRI. *NeuroImage* **6:** 122–138.

Zarahn, E., Aguirre, G. K., and D'Esposito, M. 1997. Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage* **5:** 179–197, doi:10.1006/nimg.1997.0263.