NEUROIMAGE **8,** 360–369 (1998)
ARTICLE NO. NI980369

The Variability of Human, BOLD Hemodynamic Responses

G. K. Aguirre, E. Zarahn, and M. D'Esposito

Department of Neurology, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania 19104-4283

Received November 3, 1997

Cerebral hemodynamic responses to brief periods of neural activity are delayed and dispersed in time. The specific shape of these responses is of some importance to the design and analysis of blood oxygenation level-dependent (BOLD), functional magnetic resonance imaging (fMRI) experiments. Using fMRI scanning, we examine here the characteristics and variability of hemodynamic responses from the central sulcus in human subjects during an event-related, simple reaction time task. Specifically, we determine the contribution of subject, day, and scanning session (within a day) to variability in the shape of evoked hemodynamic response. We find that while there is significant and substantial variability in the shape of responses collected across subjects, responses collected during multiple scans within a single subject are less variable. The results are discussed in terms of the impact of response variability upon sensitivity and specificity of analyses of event-related fMRI designs.

An early and consistent observation regarding blood oxygen level-dependent (BOLD), functional magnetic resonance imaging (fMRI) data is that a sudden change in neural activity produces a signal change that takes several seconds to develop and decay (Bandettini *et al.*, 1993). The sluggish nature of the BOLD fMRI signal is a consequence of its hemodynamic origins: changes in neural activity engender changes in the local vasculature and thus the local deoxyhemoglobin concentration, to which BOLD is sensitive (Malonek and Grinvald, 1996). Thus, BOLD fMRI provides a measure of the local, temporal pattern of neural activity, but only after that pattern has passed through a hemodynamic filter that smoothes and delays the signal. Because the vast majority of BOLD fMRI experiments test hypotheses regarding neural activity, as opposed to vascular physiology, methods have been developed to account for the temporal blurring imposed by the endogenous hemodynamic filter.

The particular time-course of fMRI signal change that follows a brief period of neural activity can be termed the hemodynamic response. Friston and colleagues proposed in 1994 that an estimated hemodynamic response (treated as the impulse response function of a linear system) can be used to obtain a predicted fMRI signal response for any arbitrary pattern of neural activity. These predicted signal responses can then be used to test hypotheses regarding the effect of experimental treatments upon neural activity. Because this approach more accurately predicts the shape of the fMRI signal (as compared to, for example, simply shifting a model of neural activity forward in time; Bandettini *et al.*, 1993), it affords greater statistical sensitivity and validity.

Based upon theoretical considerations, Friston and colleagues (1994) suggested that the shape of the hemodynamic response can be modeled with a Poisson function. Later, Boynton and colleagues (1996) found that a gamma function with two free parameters (plus a pure phase delay) well modeled hemodynamic responses empirically derived from the primary visual cortex of two subjects. Several groups (Courtney *et al.*, 1997; Dale and Buckner, 1997; Clark *et al.*, 1998) now use the gamma model, and the parameter fits reported by Boynton and colleagues, to generate fMRI signal predictions.

Because a single estimate of the hemodynamic response is used to analyze the data from different subjects, this approach assumes that any variability that exists between subjects in hemodynamic response is minor. If this assumption is not true, then the general approach of using a "standard" hemodynamic response to analyze BOLD fMRI data from different subjects will result in suboptimal power and perhaps invalid inference. Several groups have now anecdotally noted that observed hemodynamic responses seem to vary from subject to subject (Boynton *et al.*, 1996; Richter *et al.*, 1996; Kim *et al.*, 1997; Zarahn *et al.*, 1997b). The purpose of the present report was to formally test the hypothesis that there exists variability in hemodynamic responses collected from different subjects. In addition, we evaluated other sources that might contribute to variability seen between subjects; namely, that

1053-8119/98 $25.00

within a subject hemodynamic responses might vary from one scan to the next or from one day to the next. Measuring the relative contribution of these sources of variability in hemodynamic response can inform as to the physiological basis of response variability.

To test these hypotheses, a hemodynamic response was obtained from each of several subjects. The subjects performed a simple, event-related reaction time task in which they made a bilateral button press every 16 s during fMRI scanning. This task was assumed to produce a brief burst of neural activity within the sensorimotor strip every 16 s, and the average fMRI signal change that ensued after the button press events was taken as an estimate of the hemodynamic response for that subject. These responses were tested for the presence of significant variability. Additional subjects participated in multiple scans, either on the same or different days. The responses obtained across scans for these subjects were also tested for variability to determine if the hemodynamic response is stable within a subject across days or scans.

METHODS

MRI Technique and Initial Data Processing

Imaging was carried out on a 1.5T SIGNA scanner (GE Medical Systems) equipped with a fast gradient system for echoplanar imaging. A standard radiofrequency (RF) head coil was used with foam padding to comfortably restrict head motion. High resolution sagittal T1-weighted images were obtained in every subject. A gradient-echo, echoplanar sequence was used to acquire data sensitive to the BOLD signal at a TR = 2000 ms, TE = 50 ms. Resolution was 3.75 × 3.75 mm in plane, and 5 mm through plane, with no skip in between planes (16 or 18 axial slices acquired). A total of 160 gradient-echo echoplanar images in time were obtained per slice in each 320-s run. Twenty seconds of gradient and RF pulses preceded the actual data acquisition to allow tissue to reach steady-state magnetization.

Off-line data processing was performed on SUN Sparc workstations using programs written in Interactive Data Language (Research Systems, Boulder, CO). After image reconstruction and prior to motion correction, the data were sinc interpolated (by shifting the phase of the Fourier components) in time to correct for the interleaved fMRI acquisition sequence. This latter step is of particular importance here as hemodynamic responses were to be compared across slices that were obtained at different points in the acquisition sequence (and therefore at different points in time). If left uncorrected, this would have introduced considerable variability and bias (a phase advance) into the hemodynamic responses. The data were then motion corrected. First, a six parameter, rigid-body, least squares realignment routine was used (part of SPM96b package; Friston *et al.*, 1995b) without correction for "spin-history" (Friston *et al.*, 1996). Next, a slice-wise motion compensation method was utilized that removed spatially coherent signal changes via the application of a partial correlation method to each slice in time (Zarahn *et al.*, 1997a).

fMRI Datasets

Simple reaction time task datasets were obtained in a total of 41 (25 male) young [mean age (±SD) = 23 ± 3], right-handed subjects. Subjects viewed a backlit projection screen from within the magnet bore through a mirror mounted on the head coil. A white fixation cross was constantly illuminated in the center of a black background. Every 16 s the cross would briefly (500 ms) change to a white circle. The subject was instructed to monitor for this change, and to make a bilateral button press with both thumbs on a fiberoptic game pad. A total of 20 such trials were presented to each subject during the scan.

A 16-s intertrial interval between button press events was selected as previous reports of BOLD fMRI responses indicate that the signal change has generally run its course (i.e., has returned to baseline) after 16 s (e.g., Dale and Buckner, 1997). While residual signal changes may linger after 16 s (e.g., Boynton *et al.*, 1996), these have been reported to be rather small in magnitude and would not be expected to greatly change the shape of the average response obtained here. Nonetheless, the possibility that "overlap" of responses from one trial to the next might slightly alter the shape of the measured average response renders the hemodynamic responses reported here suspect as ideal estimates of the impulse response function of the system. It is important to note, however, that the possibility of response overlap will not weaken or bias in any way the tests of variability of responses that are the focus of this report.

Because of the regular spacing of the trials, subjects were able to anticipate the occurrence of stimuli. Such anticipation might produce increases in neural activity within the motor cortex prior to the onset of the stimulus (Georgopoulos *et al.*, 1989). This possibility also makes it difficult to treat the hemodynamic responses recorded here as good estimates of the impulse response function, as the precise moment of onset of neural activity cannot be specified.

Thirty-two of the subjects studied participated in only a single scan. One estimate of the hemodynamic response was obtained from each of these subjects. Four other subjects participated in a total of five scans; each collected on a different day, spread out over a several (5 to 11)-month period. From each of these

subjects was obtained five estimates of the hemodynamic response. Finally, five additional subjects participated in five scans all collected during a single scanning session. One of these subjects did not possess significant activity within the central sulcus during any of these scans. As a result, data from this subject could not be used in the subsequent analyses and this subject is not further discussed. From each of the remaining four subjects, five estimates of the hemodynamic response were obtained.

Creation of Statistical Maps

Voxel-wise analysis of the functional imaging data was conducted to identify voxels with a significant response to the button-press events. Statistical maps were created within the modified general linear model of Worsley and Friston (1995) using a Fourier basis set of three sines and three cosines [frequencies (Hz) = 0.0625, 0.125, 0.1875] (Josephs *et al.*, 1997). (These six covariates provided a complete basis set for the eight time points that they modeled as the data had been filtered to remove the Nyquist frequency, and nuisance covariates modeled the trial mean; see below.) Partial *F* tests were used to evaluate the significance of the variance in the data explained by these six covariates together. A specific advantage of this analysis approach is that sensitivity is not dependent on the shape of the response.

Previous work has shown that fMRI data are temporally autocorrelated under the null-hypothesis (Aguirre *et al.*, 1997a; Zarahn *et al.*, 1997a). To account for this, a mean power spectrum was obtained from each dataset by averaging the power spectra from each voxel. A 1/frequency ($1/f$) function (Zarahn *et al.*, 1997a) was then fit to this curve, ignoring those frequencies at which power attributable to task might be expected. The resulting curve was taken as an estimate of the power spectrum of the data under the null-hypothesis. The assumption underlying this analysis approach, that the noise is independent of stimulus temporal period, has been validated for fMRI (Boynton *et al.*, 1996).

The time-domain representation of the $1/f$ curve was placed within the *K* matrix (Worsley and Friston, 1995) along with a filter designed to remove low frequency confounds (below 0.025 Hz) and high frequency noise at and around the Nyquist frequency (above 0.244 Hz). It should be noted that these filtering components have no effect upon the shape of the responses obtained, as they impact frequencies that are either below that of the task or above that passed by the hemodynamic response of the system given a train of impulses as input.

Application of this analysis approach to human BOLD fMRI data collected under the null-hypothesis (similar to tests conducted in Aguirre *et al.*, 1997a, 1998; Zarahn *et al.*, 1997b) demonstrated voxel-wise false-positive rates in agreement with tabular values (data not shown).

Derivation of Hemodynamic Responses

The central sulcus was defined upon each subject's T1 images by one of the authors (GA). The central sulcus was identified as the first medial-lateral sulcus posterior to, and not in contact with, the posterior extent of the superior frontal sulcus on the superior most slices. The search volume included both the sulcus and the surrounding gray matter, yielding a total (left and right combined) search volume of ~200 voxels. The statistical maps were thresholded at an *F* value corresponding to a Bonferroni corrected, region-wise $\alpha = 0.05$, and any voxels within the search region that surpassed this threshold were identified. The average time-series from this collection of suprathreshold voxels was obtained and (i) filtered to remove frequencies below that of the paradigm and those around the Nyquist (>0.244 Hz) and (ii) adjusted to remove the effects of nuisance covariates (Friston *et al.*, 1995a). The time series was then trial averaged and adjusted to set the value of the first point to zero for display purposes. The resulting response was taken to be an estimate of the hemodynamic response of the system for that scan.

Because all voxels with a significant response within the central sulcus were averaged together, the studies conducted here are insensitive to variability in hemodynamic response from voxel to voxel within a region. Such variability might be considered to arise from differences in the diameter of the local vasculature within a voxel or the proximity of the voxel to neurally active tissue (Menon *et al.*, 1995). Future studies might seek to address whether some subset of activated voxels display more stable responses across scans within a subject than other voxels. Nonetheless, the possibility that voxel-by-voxel variability exists in the shape of the hemodynamic response within a subject will not impact the current studies that examine the variability of responses from the central sulcus region as a whole.

Statistical Analysis of IRF Variability

We considered three possible types of variability in evoked hemodynamic responses: variability (i) across subjects, (ii) within subject across days, and (iii) within subject and day across scans. To measure each type of variability, we obtained four sets of five hemodynamic responses from each category:

Responses across subjects. Of the 32 hemodynamic responses obtained from different subjects, 20 were selected at random and divided into four groups of five. This provided for four tests of the variability of this

population of responses. Because these responses were collected from different subjects, one might attribute variability in the shape of the response to the effect of subject. However, this dataset would also reflect any variability that exists between responses within the same subject across days and any variability within subject during a single scanning session.

Responses across days within subject. Four subjects were each scanned five times, each scan taking place on a different day. Each of these sets of five responses could be used to test for variability in responses obtained from the same subject on different days. While variability in response from one day to the next might explain any variability observed in the responses, variability of responses within a scanning session and subject would also be expected to contribute.

Responses across scans within subject and day. Hemodynamic responses were obtained from four subjects who each participated in five scanning runs during a single session on a single day. The source of variability in these data is scan-to-scan variability during a single session in a single subject.

Thus, four sets of five hemodynamic responses were obtained for each of three categories and tested for the presence of significant variability of response. The general approach adopted to test for variability was to use partial *F* tests to compare regression models: a full model that explicitly represented variability across hemodynamic responses with interaction terms and a reduced model that only modeled main effects of hemodynamic response. Note that in all these analyses, the hemodynamic response data (i.e., the dependent variables) were amplitude normalized to the sinc-interpolated maximum positive excursion. This was done as the hypothesized variability of interest was in the shape of the hemodynamic response (relevant for its use as a covariate) and not in its amplitude.

We wished to create a parsimonious set of independent variables for use in the aforementioned tests of variability. The remaining 12 of the 32 hemodynamic responses obtained from different subjects were examined with a principal components analysis. The first three eigenvectors of the principal components analysis were then used as covariates within regression models to test the hypotheses regarding sources of hemodynamic response variability. The hemodynamic responses from each set were concatenated and served as the dependent data. The reduced model contained three main effect covariates, one for each eigenvector, and trial effect covariates to mean center the residuals of each response. The full model was identical to the reduced model except for the addition of covariates which modeled interactions between the hypothesized variability source and the behavior of the hemodynamic response. In these interaction covariates, each of the three chosen eigenvectors was modeled separately. A graphical example of the design matrices corresponding to a reduced and a full model is provided in Fig. 1. The results of each hypothesis test were assessed by performing an appropriate partial *F* test comparing the full and reduced models (Kirk, 1982, p. 179). Validation of the model was performed using null-hypothesis simulations in which five identical hemodynamic responses (plus normally distributed, computer generated noise) served as the dependent data. These tests confirmed that the test yields tabular false-positive rates under the null-hypothesis.

The results of these tests of variability were 12 *F* values, 4 from each category (i.e., across subjects, within subject across days, and within subject and session across scans). An additional analysis determined if significantly greater variability (i.e., greater *F* values) was observed in one group versus another. Because of the nonnormal distribution of the dependent data (i.e., *F* values) nonparametric tests were employed. All 12 values were entered into a Kruskal–Wallis analysis to determine if a significant effect of category was present within the data. Then, pair-wise Mann–Whitney tests were used to determine if greater *F* values were present within one given category as compared to another.

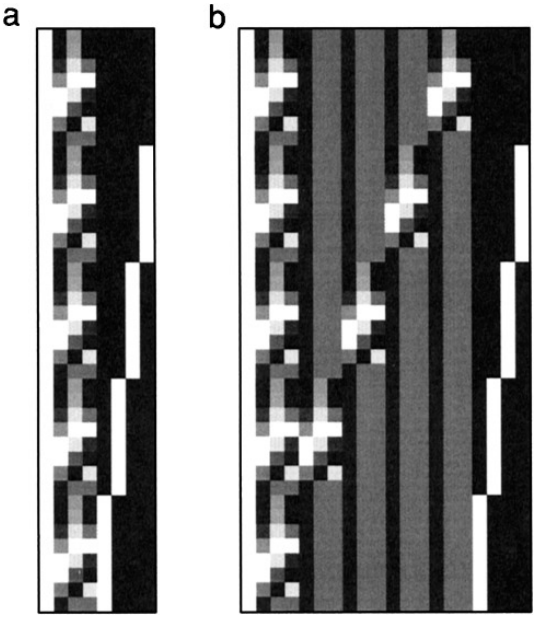

FIG. 1. Example design matrices used to test for hemodynamic response variability. (a) The reduced model. Arranged in columns from left to right, the model contains an overall intercept, three main effect covariates, and four trial-effect covariates. Each main effect covariate is composed of multiple, shifted versions of one of the eigenvectors derived from the central sulcus hemodynamic responses from 12 subjects (see Fig. 3 and Table 1). (b) The full model. This model is identical to that shown in a, except for the addition of interaction terms designed to model the variability across hemodynamic responses. These two models were compared with a partial *F* test (see Methods).

RESULTS

IRF Descriptive Properties

Voxels that evidenced significant signal changes in response to the button-press task were found within the central sulci of all 32 of the subjects scanned once [mean number of voxels (±SD) within central sulcus region across subjects: 31 ± 20, range: 6–91]. Across subjects, the mean (±SD) peak signal change from baseline was 2.05% ± 0.53 (n = 32, range: 1.0 to 3.1%). It is not possible to attribute a particular source to this variability in response amplitude as it could be due to differences in the amplitude of underlying neural activity (perhaps related to differences in amplitude of motor response), differences in the transformation properties of the system, variations in nonphysiological properties of the MRI scanner, or any combination of these. As we are interested here in variability in the filtering properties of the hemodynamic responses (i.e., the shape of the response), absent differences in raw amplitude, all hemodynamic responses analyzed and presented here have been normalized to the maximum positive excursion.

Twenty of the 32 hemodynamic responses obtained from different subjects displayed some degree of delayed undershoot (i.e., signal values fell below baseline postpeak), although there was some variability as to its degree [mean undershoot as a proportion of maximum positive excursion (±SD): −0.29 ± 0.22 (n = 20; range: −0.02 to −0.81)]. It thus seems that models of the hemodynamic response that do not include an undershoot (Friston *et al.*, 1994; Boynton *et al.*, 1996) will be unable to completely represent these empirically derived responses.

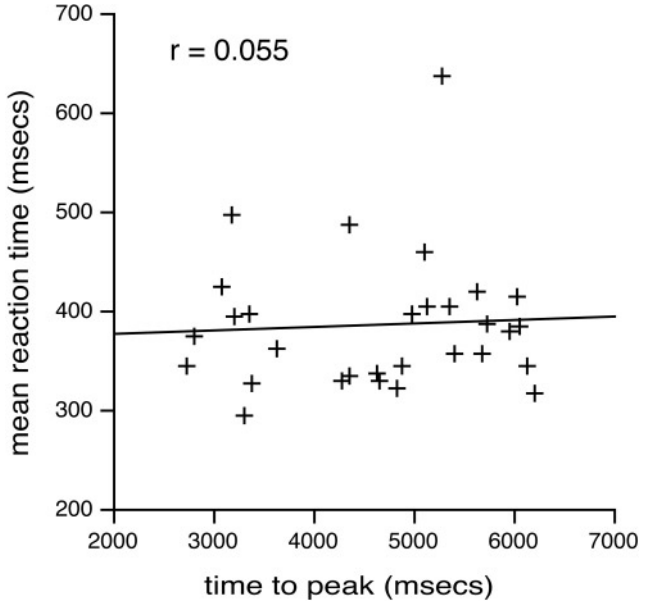

FIG. 2. Mean reaction time for each of 30 subjects versus the time to peak of the hemodynamic response obtained from that subject's central sulcus. No significant relationship between the measures is present.

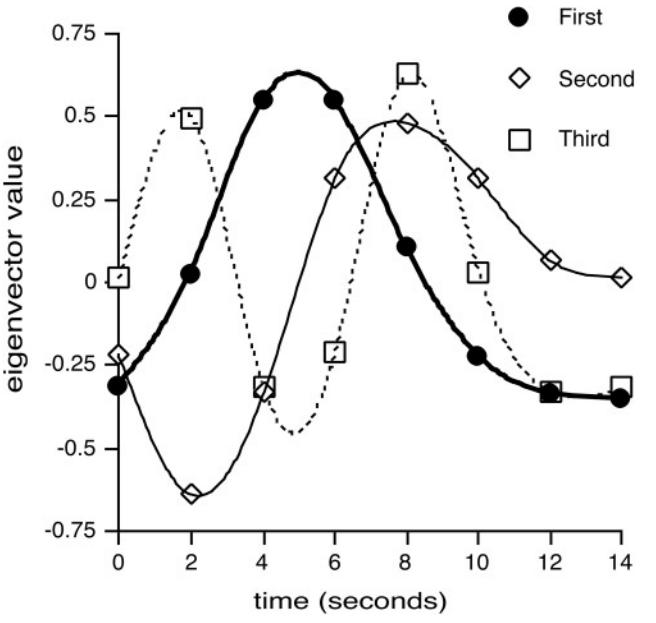

FIG. 3. First three eigenvectors from a principle components analysis of central sulcus hemodynamic responses from 12 different subjects. All vectors are orthonormal. The values of the eigenvectors and their corresponding eigenvalues are provided in Table 1.

The sinc-interpolated time-to-peak was obtained for all 32 hemodynamic responses. The mean (±SD) time-to-peak of 4.7 ± 1.1 s (n = 32, range: 2.7 to 6.2) is in rough agreement with that observed in previous studies of the BOLD hemodynamic system (Boynton *et al.*, 1996; Richter *et al.*, 1996; Kim *et al.*, 1997). Behavioral (reaction time) data were obtained for 30 of the 32 subjects [mean RT (±SD) = 368 ± 68 ms]. There was no discernible relationship between a subject's time-to-peak and reaction time [r (28 df) = 0.055, NS] (see Fig. 2) in agreement with previous studies (Richter *et al.*, 1996; Kim *et al.*, 1997). While we might expect a linear, unit slope relationship between these measures, our failure to observe a significant correlation is unsurprising given the much larger variance in hemodynamic time-to-peak relative to reaction time variance.

Generation of Eigenvectors of Evoked Hemodynamic Responses

Twelve of the hemodynamic responses obtained from subjects scanned a single time were selected at random. These responses were subjected to a principle components analysis. Figure 3 presents the first three eigenvectors from that analysis, which together explained >98% of the variance within the set. The specific values of the three vectors and their eigenvalues are provided in Table 1. As might be expected, the first principle

component strongly resembles the shape of hemodynamic responses previously reported by several groups (Boynton *et al.*, 1996; Richter *et al.*, 1996). These eigenvectors are used below in the analysis of variability of hemodynamic responses. We also note that these three vectors may serve as a parsimonious set of covariates in the analysis of other event-related fMRI experiments.

Tests of IRF Variability

Several groups have anecdotally noted that hemodynamic responses seem to differ from subject to subject (e.g., Boynton *et al.*, 1996; Kim *et al.*, 1997) and that responses appear to be more stable during a single scanning session with a single subject (Kim *et al.*, 1997). We wished to test these observations and to measure the variability observed in responses collected within subjects on a single day, within subjects across days, and across subjects.

We asked, first, whether significant variability exists among hemodynamic responses collected from different subjects. Twenty hemodynamic responses obtained from 20 different subjects were randomly divided into four groups of five. The purpose of this division was to permit multiple tests of variability upon independent datasets, thus allowing for replication. Each group of responses was tested for the presence of significant variability. Variability was detected using partial F tests to compare statistical models that either explicitly modeled the presence of variability among the responses or did not. Both models contained three covariates (generated using the three eigenvectors) that modeled the "main effect" of responses across subjects (see Fig. 1a). In effect, these covariates modeled the mean response, or the components of the hemodynamic responses that were shared across subjects within the group. The full model, but not the reduced model, also contained interaction terms. These covariates modeled, for each response, aspects of the response that were not shared across the group of five. The statistical test measured if a significantly greater proportion of variance was explained by these interaction covariates than would be expected by chance alone.

TABLE 1

Values of the Eigenvectors Shown in Fig. 2

	Principal components		
Time (s)	First	Second	Third
		Eigenvectors	
0	−0.316	−0.217	0.017
2	0.025	−0.641	0.498
4	0.547	−0.331	−0.316
6	0.551	0.317	−0.211
8	0.104	0.480	0.629
10	−0.222	0.311	0.032
12	−0.335	0.067	−0.331
14	−0.353	0.013	−0.319
		Eigenvalues	
	0.743	0.222	0.018

Note. The eigenvalues have been normalized to the total variance.

TABLE 2

Results of Partial F Tests of Variability amongst Hemodynamic Responses

	Partial F tests (12, 20 df)		
	Across subject	Across day	Across scans
Set 1	**17.6**	**2.3**	1.7
Set 2	**20.2**	**9.6**	0.6
Set 3	**7.8**	**5.7**	**2.5**
Set 4	**26.9**	1.9	0.4
Median	18.9	4	1.15

Note. Four tests were conducted for each category of variability. Significant ($P < 0.05$) responses in bold.

A partial F test revealed that the subject-by-response interaction covariates explained a significant amount of variance for all four sets of responses (see Table 2). Thus, there are significant differences between hemodynamic responses collected across subjects from the same anatomical region. This test does not reveal, however, whether or not this variability can be attributed to the effect of varying subject, day, and/or scan session (as these are all confounded). Figure 4a presents the set of across-subjects responses with the least variability while Fig. 4b presents the set with the greatest variability.

Responses were also collected across multiple days from single subjects. In total, four subjects were each scanned five times, with each scan occurring on a different day over the course of several months. If significant differences exist between these responses in a single subject, then scanning on different days can contribute to response variability. Three of the four subjects were found to have responses that varied significantly from one day to another. However, and again, day was confounded here with "scan" as these responses were obtained during different scanning sessions. Figures 4c and 4d present two of these sets of responses.

Finally, five hemodynamic responses were collected from each of four subjects during a single scanning session on a single day. The existence of significant differences between the responses obtained from a single subject in this setting would suggest that hemodynamic response variability exists from scan to scan. Partial F tests found significant variability in only one

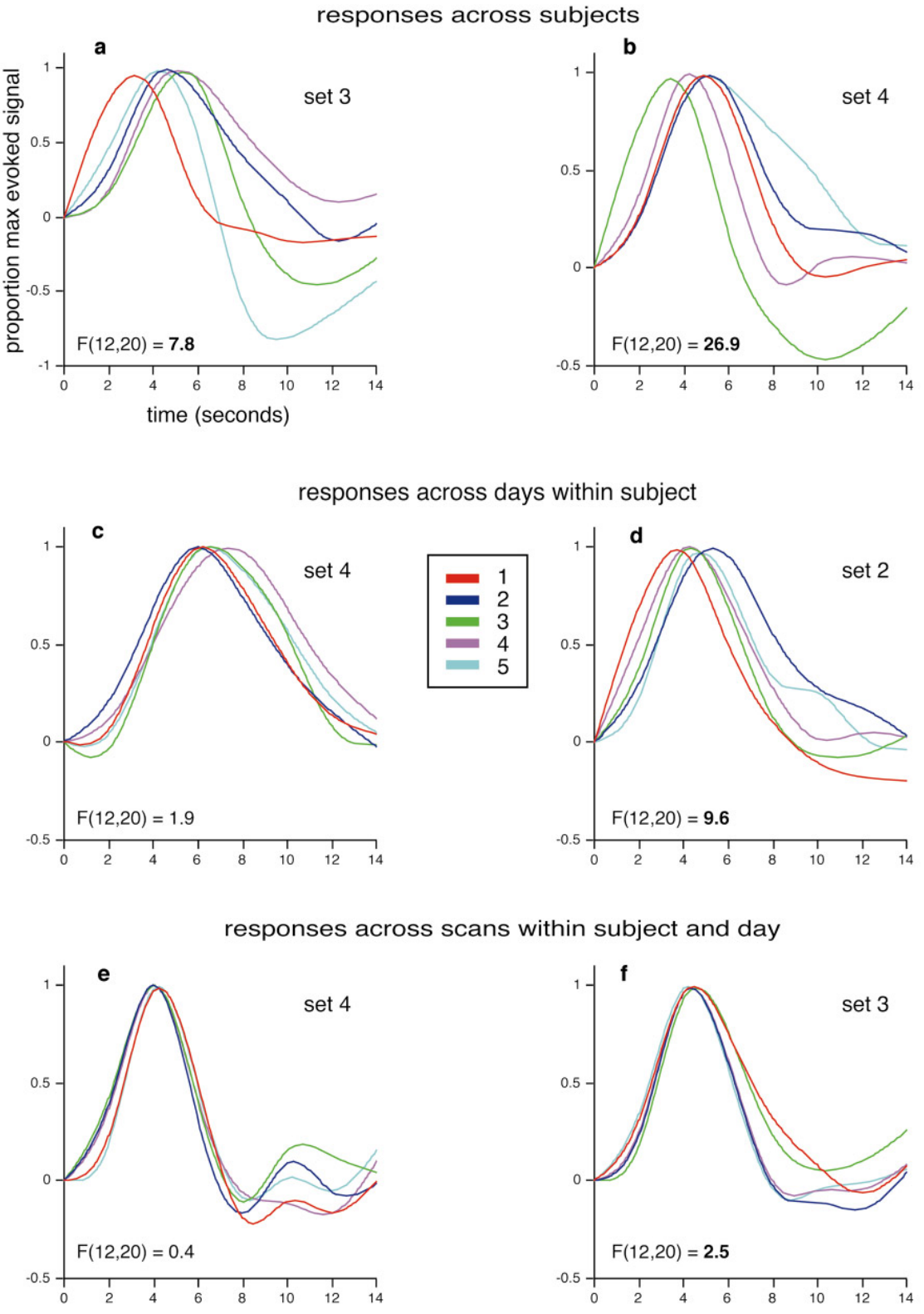

FIG. 4. Hemodynamic responses from the central sulcus, each sinc-interpolated and normalized to the maximum, interpolated positive excursion. (a, b) Responses obtained from two different groups of five subjects each, representing the groups with (respectively) the least and greatest variability in response shape. (c, d) Responses (least and greatest variability) collected from each of two subjects across different days. (e, f) Responses (least and greatest variability) collected from each of two subjects across multiple scans in a single day. The *F* values reported for each set of responses are the results of tests of variability (see Methods and Results). The key indicates the order of acquisition of the responses for those collected within subjects (c–f).

of the four subjects studied. Figures 4e and 4f present two of these sets of responses from within subject and day.

As can be seen from the median F scores (Table 2) for each category of responses studied, the variability in responses collected on different days is roughly 3.5 times greater than the variability present in responses collected during a single scanning session. Moreover, the median variability observed across subjects is over 16 times greater than that observed within a scanning session. A Kruskal–Wallis test performed upon the F values collected confirmed that there was a significant effect of category (i.e., across subjects, across days, across scans) upon the variability of responses observed ($P = 0.02$). Subsequent pair-wise tests revealed that responses across subjects were significantly more variable than either responses within a scanning session (Mann–Whitney rank test, $P = 0.02$) or responses across days (Mann–Whitney rank test, $P = 0.04$). Responses across days, however, were not found to be significantly more variable than responses within a scanning session (Mann–Whitney rank test, $P = 0.08$). Finally, it should be noted that differences in variability of mean RT across subjects compared to within subject cannot account for these results, given the absence of any correlation between RT and time-to-peak observed above.

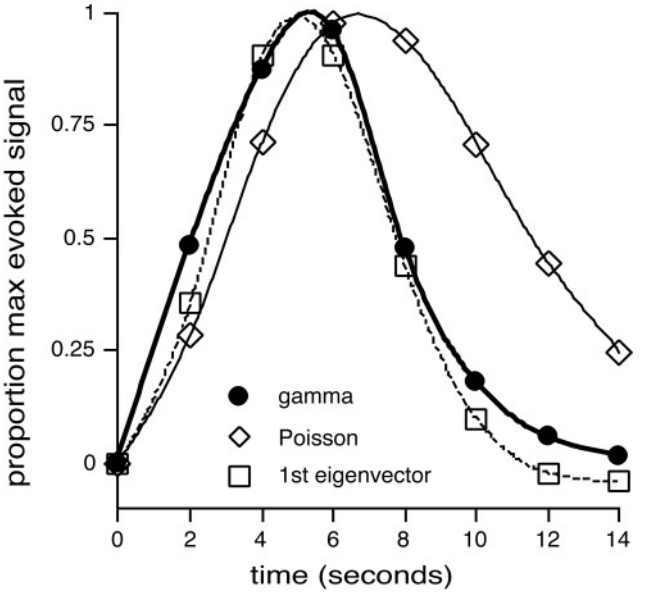

FIG. 5. Three different summary models of the hemodynamic response function. The thick solid line (circles) is the vector described by the gamma function of Boynton and colleagues (1996) [parameters: $\tau = 1.25$, $n = 3$, $\delta = 2.5$]. The thin solid line (diamonds) is a Poisson function [parameter = 8 s] deconvolved with a Gaussian, $\sigma =$ 1.9 s, to account for processing performed by Friston and colleagues (1994). The dotted line (squares) is the first eigenvector from Fig. 3.

Response Variability and Standard Models of the Hemodynamic Response

The significant variability in the shape of the hemodynamic response function across subjects calls into question the common use of a single, representative response function to model fMRI data from all subjects. Here we examined how well standard response functions fit a population of responses from different subjects. We then tested if an alternative approach might give rise to greater sensitivity: Given the stability of the hemodynamic response observed within a single subject, a promising analysis approach is to empirically derive a hemodynamic response for every subject, and then use that subject-specific response to analyze further BOLD fMRI data from that subject.

We examined three models of the hemodynamic response (see Fig. 5). The first two of these are routinely used for the generation of covariates for the analysis of BOLD fMRI data: the Poisson function described by Friston and colleagues (1994) and the gamma function described by Boynton and colleagues (1996). The third model examined was the first eigenvector presented here (Fig. 3). We measured the correlation between each of these models and each of the 20 hemodynamic responses obtained from different subjects (recall that the first eigenvector was generated from an independent data set). Of interest was the degree to which one model or another tended to produce higher correlation values, indicating a better fit between the model and the population of responses. T tests performed upon the t-transformed correlation values were used to assess the significance of these differences.

The Poisson model had a mean ($\pm$SD) correlation with the empirically obtained responses of 0.490 ± 0.390 ($n = 20$; range: -0.166 to 0.917). This indicates that the Poisson function, on average, explained only 25% of the variance present in the evoked responses. In contrast, the gamma model, and the first eigenvector reported here, both explained nearly 70% of the variance in the evoked response on average. The average ($\pm$SD) correlation of the gamma model with the population of responses was 0.826 ± 0.147 ($n = 20$; range = 0.509 to 0.976) and the average for the first eigenvector was 0.833 ± 0.152 ($n = 20$; range = 0.501 to 0.991). The correlation values produced by both of these models were significantly higher than those of the Poisson function [$t(19\ df) = 4.54, 3.87$, $P < 0.001$].

Included in our datasets are four subjects who each were scanned five times during a single scanning session, producing five hemodynamic responses per subject. We correlated the first hemodynamic response from each subject with the subsequent responses from that subject, resulting in four correlation values per subject. The mean correlation value from each subject was then taken to provide a set of four representative correlation values, one from each subject, that describe the correspondence between a subject-specific hemody-

namic response and subsequent responses from that subject during a scanning session. The mean (±SD) of these correlation values was 0.964 ± 0.12 ($n = 4$; range = 0.947 to 0.976), indicating that using a subject-specific response function results in a model that can explain, on average, 92% of the variance in subsequent evoked responses. The correlation values obtained from this approach were significantly higher than those obtained from even the best of the standard models tested just above [$t(22\ df) = 2.27$, $P = 0.03$]. The additional explanatory power of this approach is not only significant, but substantial as well—an additional 22% of the variance of the hemodynamic response, on average, was explained by using a subject-specific hemodynamic response.

DISCUSSION

Given the stability of the hemodynamic response within subject relative to that across subjects, it seems that the particular shape of the hemodynamic response within the central sulcal region is largely a consequence of intersubject variability in physiology. The smaller variability in responses within subject across days suggests that this response may change over time for some subjects. However, because subjects were not studied during different scanning sessions on the same day (i.e., removed from the scanner and then repositioned), it is not possible to unambiguously attribute the variability in response across days to the effect of day per se, as it is conceivable that different scanning sessions alone might introduce variability.

As has been noted, several groups make use of *a priori* estimates of the hemodynamic response of the BOLD fMRI system to test functional neuroimaging hypotheses. This type of analysis approach can be contrasted with alternative designs that use a flexible basis set instead. These approaches, like the Fourier basis set design used here (Josephs *et al.*, 1997), are notable for their equivalent sensitivity to any consistent pattern of signal change associated with neural events of interest. While the reduced assumptions inherent in such an analysis approach are sometimes desirable, the lack of a model for the hemodynamic transform also limits the nature of inferences that can be drawn. This is because "significant" signal changes might be due to any one of numerous differences between two experimental conditions. Thus, *a priori* estimates of the hemodynamic response will continue to be necessary when focused hypotheses regarding the timing or intensity of neural activity are to be tested.

The presence of response variability across subjects calls into question the use of a single representation of a hemodynamic response to model evoked BOLD fMRI signal changes in different subjects. Specifically, covariates generated using a standard model will not be perfectly valid, in that they will be unable to completely model experimentally introduced variance. The particular impact of this error, as well as its degree, can be expected to vary greatly from one experimental design to another (Aguirre and D'Esposito, 1998).

For traditional "blocked" fMRI designs, in which relatively long duration periods of neural activity are evoked, failure to perfectly model the shape of the hemodynamic response will likely result in only small decrements in sensitivity. Nonetheless, the use of a more accurate response function can result in measurable increases in sensitivity, as has been demonstrated for analyses of blocked fMRI experiments that used either a Poisson function or an empirically derived estimate of the hemodynamic response (Aguirre *et al.*, 1997a).

For "event-related" fMRI designs, accurate estimation of the hemodynamic response becomes more important. For designs in which different trial types are randomly ordered (or perfectly counterbalanced) (e.g., Dale and Buckner, 1997; Clark *et al.*, 1998), failure to accurately specify the shape of the hemodynamic response can be expected to result in decrements in sensitivity. This loss of power will likely increase as the spacing between the trials decreases (as the task variance moves to ever higher frequencies) and might be substantial.

It should be noted, however, that inaccurate characterization of the hemodynamic response will not lead to invalid inference under these circumstances. This is because the random ordering of the trials makes any difference between the assumed and actual hemodynamic response unbiased. This is to be contrasted with event-related designs in which the different events of interest cannot be randomly ordered—classically, working-memory paradigms in which a delay period always follows the presentation of a stimulus to be remembered (Courtney *et al.*, 1997; Zarahn *et al.*, 1997b). For designs of this kind, it is essential that the signal change produced by one event type (e.g., the stimulus presentation), be well modeled and not erroneously attributed to an effect of another event type (e.g., the delay period). Inferential false-positive results would be expected if, for example, the actual hemodynamic response of the subject was broader in time than predicted by the assumed hemodynamic response. The distribution of time-to-peak values in responses obtained across subjects here (range: 2.7 to 6.2 s) suggests that this type of error of inference is a real possibility and should be guarded against by additional control conditions and tests in event-related designs of this kind (Zarahn *et al.*, 1997b).

These considerations lead us to advocate the use of subject-specific hemodynamic responses for the analysis of BOLD fMRI data. There is, however, an important caveat to this suggestion. All of the responses that

were studied here were obtained from within the central sulcus. Anecdotal evidence suggests that the shape of the hemodynamic response varies from one cortical region to another, although proof of this assertion remains elusive, as it is unknown whether changes in neural activity or vascular responsiveness underlie the regional response differences. Thus, the application of a hemodynamic response derived from the primary motor cortex to test hypotheses in other areas of the brain may also be strictly invalid. Future studies will be necessary to determine if (i) the hemodynamic response actually does vary from region to region and (ii) if so, if this variability from region to region is greater than the variability within a region across subjects. Alternatively, one might attempt to obtain an estimate of the hemodynamic response from each cortical area under study during preliminary studies.

Finally, the stability of the response observed here within subject suggests the feasibility of fMRI designs that attempt to detect small neural onset asynchronies. Because the variability in the time-to-peak of hemodynamic responses collected from a single subject on a single day is rather minor, small (e.g., 50–100 ms) differences between two event types in the onset of neural activity in a single region may be detected over the course of repeated trials.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (NS01762 and AG13483) and the Charles A. Dana Foundation and the American Federation for Aging Research. We also thank Jessica Lease and Rajiv Singh for their help in data collection.

REFERENCES

Aguirre, G. K., and D'Esposito, M. 1998. Experimental Design for Brain fMRI. In *Functional MRI* (P. A. Bandettini and C. Moonen, Eds.), in press. Springer Verlag, Berlin.

Aguirre, G. K., Zarahn, E., and D'Esposito, M. 1998. A critique of the use of the Kolmogorov-Smirnov (KS) statistic for the analysis of BOLD fMRI data. *Magn. Reson. Med.* **39:**500–505.

Aguirre, G. K., Zarahn, E., and D'Esposito, M. 1997a. Empirical analyses of BOLD fMRI statistics. II. Spatially smoothed data collected under null-hypothesis and experimental conditions. *Neuroimage* **5:**199–212.

Bandettini, P. A., Jesmanowitz, A., Wong, E. C., Hyde, J. S. 1993. Processing strategies for time-course data sets in functional MRI of the human brain. *Magn. Reson. Med.* **30:**161–173.

Boynton, G., Engel, S., Glover, G., Heeger, D. 1996. Linear systems analysis of functional magnetic resonance imaging in human Vl. *J. Neurosci.* **16:**4207–4221.

Clark, V. P., Maisog, J. M., Haxby, J. V. 1998. An fMRI study of face perception and memory using random stimulus sequences. *J. Neurophysiol.*, **79:**3257–65.

Courtney, S. M., Ungerleider, L. G., Keil, K., Haxby, J. V. 1997. Transient and sustained activity in a distributed neural system for human working memory. *Nature* **386:**608–611.

Dale, A. M., Buckner, R. L. 1997. Selective averaging of rapidly presented individual trials using fMRI. *Hum. Brain Map.* **5:**329–340.

Friston, K., Ashburner, J., Frith, C., Poline, J.-B., Heather, J., Frackowiak, R. 1995b. Spatial registration and normalization of images. *Hum. Brain Map.* **2:**165–189.

Friston, K., Williams, S., Howard, R., Frackowiak, R., Turner, R. 1996. Movement-related effects in fMRI time-series. *Magn. Reson. Med.* **35:**346–355.

Friston, K. J., Holmes, A. P., Poline, J.-B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J., Turner, R. 1995a. Analysis of fMRI time-series revisited. *NeuroImage* **2:**45–53.

Friston, K. J., Jezzard, P., Turner, R. 1994. Analysis of functional MRI time-series. *Hum. Brain Map.* **1:**153–171.

Georgopoulos, A. P., Crutcher, M. D., Schwartz, A. B. 1989. Cognitive spatial motor processes: 3. Motor cortical prediction of movement direction during an instructed delay period. *Exp. Brain Res.* **75:**183–194.

Josephs, O., Rees, G., Turner, R., Friston, K. J. 1997. Event-related fMRI. *Hum. Brain Map.* **5:**243–248.

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

Kirk, R. E. 1982. *Experimental Design: Procedures for the Behavioral Sciences.* Brooks/Cole, Monterey, CA.

Malonek, D., and Grinvald, A. 1996. Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: Implications for functional brain mapping. *Science* **272:**551–554.

Menon, R. S., Ogawa, S., Hu, X., Strupp, J. P., Anderson, P., and Ugurbil, K. 1995. BOLD based functional MRI at 4 tesla includes a capillary bed contribution: Echo-planar imaging correlates with previous optical imaging using intrinsic signals. *Magn. Reson. Med.* **33:**453–459.

Richter, W., Ugurbil, K., and Kim, S.-G. 1996. Limitations of temporal resolution in fMRI. *NeuroImage* **3:**S38.

Worsley, K. J., and Friston, K. J. 1995. Analysis of fMRI time-series revisited—Again. *Neuroimage* **2:**173–182.

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

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