NEUROIMAGE **5,** 179–197 (1997)
ARTICLE NO. NI970263

Empirical Analyses of BOLD fMRI Statistics

I. Spatially Unsmoothed Data Collected under Null-Hypothesis Conditions

E. Zarahn, G. K. Aguirre, and M. D'Esposito[1]

Department of Neurology, University of Pennsylvania Medical Center, Philadelphia, Pennsylvania 19104-4283

Received September 9, 1996

Temporal autocorrelation, spatial coherency, and their effects on voxel-wise parametric statistics were examined in BOLD fMRI null-hypothesis, or "noise," datasets. Seventeen normal, young subjects were scanned using BOLD fMRI while not performing any time-locked experimental behavior. Temporal autocorrelation in these datasets was described well by a 1/frequency relationship. Voxel-wise statistical analysis of these noise datasets which assumed independence (i.e., ignored temporal autocorrelation) rejected the null hypothesis at a higher rate than specified by the nominal α. Temporal smoothing in conjunction with the use of a modified general linear model (Worsley and Friston, 1995, *NeuroImage* 2: 173–182) brought the false-positive rate closer to the nominal α. It was also found that the noise fMRI datasets contain spatially coherent time signals. This observed spatial coherence could not be fully explained by a continuously differentiable spatial autocovariance function and was much greater for lower temporal frequencies. Its presence made voxel-wise test statistics in a given noise dataset dependent, and thus shifted their distributions to the right or left of 0. Inclusion of a "global signal" covariate in the general linear model reduced this dependence and consequently stabilized (i.e., reduced the variance of) dataset false-positive rates.

INTRODUCTION

The detection of changes in neural activity using blood oxygenation level-dependent functional magnetic resonance imaging (BOLD fMRI; Kwong *et al.,* 1992; Ogawa *et al.,* 1993) generally involves the identification of voxel signals which correlate with an imposed experimental paradigm (Bandettini *et al.,* 1993). Theoretical null-hypothesis probability distributions are used to determine the probability of obtaining a given or greater relationship due to chance alone. A valid statistical model is necessary for the actual expected false-positive rate for an experiment (also referred to as type I error or false alarms) to equal the theoretically expected false-positive rate. The validity of a statistical model depends upon the extent to which the data in question satisfy the model's assumptions. Typically, fMRI data are analyzed with some variant of the general linear model (GLM); i.e., correlations, *t* tests, or multiple regression (Friston *et al.,* 1995d). An important assumption of the GLM is the independence of observations. When this assumption is violated, estimates of parameter errors can be inappropriate, leading to expected false-positive rates not in agreement with theory (Worsley and Friston, 1995). Observations have been made that suggest that fMRI data are temporally autocorrelated (or nonindependent in time; Weisskoff *et al.,* 1993; Friston *et al.,* 1994; Boynton *et al.,* 1996), thus calling into question the validity of the GLM for fMRI data. Our first aim was thus to characterize the intrinsic autocorrelations in fMRI time-series data. Using gradient echo–echoplanar fMRI, we collected "noise" datasets in which subjects were neither required to perform any experimentally time-locked behavior nor exposed to any time-varying experimental stimuli. If the fMRI data are independent in time, we would expect the spatially averaged power spectrum of fMRI time-series to be flat. The power spectra of the noise datasets were thus used to test for autocorrelation by looking for consistent characteristics of their structure. Two specific models (a 1/frequency model and a decaying exponential model) were compared for their ability to describe the noise power spectra. Data from water phantoms were also collected to allow (by comparison with the human datasets) consideration of physiological mechanisms of autocorrelation.

Worsley and Friston (1995) have provided a modified GLM in which the assumed autocorrelation structure of the data can be incorporated. In their analysis, they assumed an autocorrelation structure determined by a Poisson filtering of the data. It has been previously suggested that intrinsic temporal autocorrelation (i.e., autocorrelation present in the raw data before exog-

[1] To whom reprint requests should be addressed.

1053-8119/97 $25.00

enous temporal filtering is applied in processing) might be negligible compared to the autocorrelations introduced after smoothing (Friston *et al.*, 1995c). We wished to test this assumption empirically. We also wished to see if the satisfaction of this assumption depended on the specific smoothing filter used. The noise datasets were analyzed using the modified GLM with differing models of intrinsic autocorrelation and two types of temporal smoothing kernels (one band slightly broader than the other). In these analyses, we assumed a periodic "on–off" behavioral paradigm. Since no time-locked task was performed, the null hypothesis should be true. The observed false-positive rates for the test statistic corresponding to the assumed task parameter obtained from the various versions of the modified GLM analyses were compared to the expected false-positive rate derived from theoretical test statistic distributions.

Another question regarding fMRI analysis is the possibility of spatial coherency (i.e., the presence of cross-correlations of voxel time series) in fMRI noise datasets which have not been spatially smoothed during processing. Spatial coherency of voxel time series is an important issue because certain statistical methods make assumptions about independence of the dependent variables under the null hypothesis. One type of spatial coherency that has been observed in neuroimaging data is typically referred to in positron emission tomography (PET) as "global flow" (Friston *et al.*, 1990), which is simply the average measured blood flow for the entire volume. In fMRI there is at least a superficial analogue which we will refer to as the "global signal" and which we define as the average brain voxel time series (the global signal is thus a time signal). This global signal may be used to assess if the voxel time series are, to some extent, spatially coherent. Any such observed spatial coherency could be due to a stationary, continuously differentiable spatial autocovariance function or to other types of spatial smoothness. The former type of spatial smoothness is relevant in the context of statistical parametric mapping techniques (Friston *et al.*, 1991; Worsley *et al.*, 1992). Both types of spatial smoothness were examined in this report.

It has been common practice to include a global covariate in the analysis of PET studies (Friston *et al.*, 1990). Some fMRI studies (Friston *et al.*, 1995b) have included a global signal covariate as well. A possible advantage to this method is that it may reduce spatially coherent noise in the data, thus increasing power to detect activations. The effect of this practice would depend, in part, on the temporal and spatial nature of the global signal in data in which the null hypothesis was true. We therefore examined the spatial coherency of the global signal. The effect of accounting for the global signal on the specificity of the analysis of the noise data was also directly examined by comparing the GLM results obtained with and without a global signal covariate in the noise datasets. Additionally, the gross spatial localization of the global signal was examined to see if it segregated obviously to a particular tissue type (gray matter, white matter, or ventricle). This information is relevant in the context of attempting to determine if there are any physiological mechanisms contributing to spatial coherency, as well as in considering the possible functional effects of including a global signal covariate in analysis.

METHODS

MRI Technique

Imaging was carried out on a 1.5-T SIGNA scanner (GE Medical Systems) equipped with a prototype fast gradient system for echo-planar 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 T_R = 2000 ms, T_E = 50 ms. Resolution was 3.75 × 3.75 mm in plane and 5 mm between planes. The number of axial slices and images in time per slice are provided for each dataset. Twenty seconds of "dummy" gradient and RF pulses preceded the actual data acquisition.

Motion Compensation

A slice-wise motion correction method was utilized which removed spatially coherent signal changes via the application of a partial correlation method to each slice in time. For each axial slice at each time, a difference image between that slice at Time *t* and that slice at Time 0 (a Motion image) was correlated with an image composed of the difference between the slice at Time 0 shifted to the right 1 voxel and that same slice shifted to the left 1 voxel (an X-shift image). The same operation was performed for *y* shifts (using Y-shift images). The X-shift and Y-shift images, weighted by the strength of their respective correlations with the Motion image, were subtracted from the image of the slice at Time *t*. Thus, the rationale of this method was to subtract out signal changes that correlated with small (on the order of a voxel) translations in the *x* and *y* dimensions. A conceptually similar method for motion in the *z* dimension was then applied to each axial image. A pseudo-Z-shift image was computed for each axial slice by subtracting the average of the first 10 images (in time) from the average of the last 10 images. The Z-shift image, weighted by its correlation with a Motion image (computed after applying the corrections described above for *x* and *y* shifts) was then subtracted from the slice image at Time *t*. The rationale for the

pseudo-Z-shift image was that translation in the *z* dimension would occur steadily throughout the acquisition of images within a run.

Although this method does not take rotations into account and does not compensate for *x*, *y*, and *z* motion simultaneously, it is a rapid method that does reduce the voxel variance and the appearance of artifacts (unpublished observations). Of course, as it does not actually correct for motion per se (in that it does not actually solve for any transformation back to a reference image, but rather compensates for signal changes that correlate with movement, a subtle but important difference), it is less desirable in theory than a true motion-correction algorithm.

Primary Group of Human Noise Datasets

Noise datasets were obtained by scanning 17 healthy, young (22–34 years of age, mean = 25; 12 M and 5 F) subjects. Sixteen axial slices and 160 images/slice were acquired per subject. Subjects were instructed to rest with eyes open (to otherwise mimic experimental conditions and to discourage sleep). The room was dim, but lights from the control room were visible. Equipment used in our laboratory to present stimuli during behavioral imaging experiments [including an overhead projector plugged in inside the scanning room and an LCD panel (InFocus systems panelbook 550) in the room connected to a Macintosh 280c computer outside the room] were kept on and in their standard setups to include all sources of noise typical to our experiment (see below). Characterization of intrinsic temporal autocorrelation, examination of false-positive rates, and examination of global signals were performed upon these datasets as described.

Estimation of the BOLD Impulse Response Function

A separate behavioral dataset (Shin *et al.*, 1995) was used only for the purposes of determining the putative impulse-response function (IRF) of the system mediating the transformation of neural activity to BOLD signal. This dataset comprised a separate group of 12 subjects who performed a visuospatial discrimination task that alternated with a sensorimotor control at 0.0125 Hz. The task was adapted from a study presented by Haxby and colleagues (1991). Subjects performed one visuospatial discrimination every 5 s during the experimental condition or a left–right alternating button press every 5 s in the sensorimotor control condition (for more details of the task, see companion paper Aguirre *et al.*, 1997, and Shin *et al.*, 1995). The scanning parameters were exactly the same as those for the 17 human "noise" datasets. Using this behavioral dataset, the signals from all parietal lobe (i.e., posterior to central sulcus, superior to lateral occipital sulcus, and sylvian fissure) voxels of a cluster size of 2 or more and with $R > 0.2$ (cluster size and correlation threshold were arbitrarily chosen) were averaged together to yield a group average activation signal. No other results from this dataset will be presented here (for other results see Shin *et al.*, 1995).

Assuming a step function of neural input and linearity and time invariance of the transformation of change in neural activity to BOLD signal change, an estimate of the BOLD IRF was obtained by subtracting the group average activation signal shifted to the right 1 time point from the unshifted group average activation signal. This method is based upon the invertibility of the step transfer function, and essentially performs a first-differencing operation on the observed discrete time output signal (Oppenheim *et al.*, 1983). This method is to be contrasted with that of Friston and colleagues (1994). In that report, the method used to estimate the BOLD IRF did not rely on task effects, but rather depended upon assumptions about the source of correlated noise components in fMRI time series.

Conceptual Basis of Estimation of Intrinsic Temporal Autocorrelation

The basic concepts used here to investigate temporal autocorrelations will be briefly described. A periodic discrete time signal can also be expressed as a discrete signal in frequency with the same number of data points and with no loss of information. The theory and method behind this transformation is called the Discrete Fourier Transform (or DFT; Oppenheim *et al.*, 1983). The squared magnitudes of the DFT vs frequency are called the power spectrum. The power spectrum function is a frequency domain representation of the autocorrelation function. Parseval's relation for discrete, periodic signals is a theorem that relates the variance of a signal to the sum of its power spectrum across frequency,

$$\frac{1}{N}\sum_{n=1}^{N}|x_n|^2 = \sum_{k=1}^{N}|a_k|^2,$$

where the x_n are the values of the discrete (time) signal at time n, the a_k are the DFT coefficients at frequency k, and the sum is over one period of the signal in time or over the harmonics of the fundamental frequency. Parseval's relation was used here to characterize the proportion of the variance (or power) of fMRI time series attributable to certain frequencies. In uncorrelated (or white) noise, the expected power is evenly distributed across all frequencies. If there is a systematically uneven distribution of power across frequencies, then the noise is autocorrelated (or colored).

Voxel-Averaged Power Spectra Determination

For this and all subsequent analyses, "brain voxels" were defined as those with a minimum signal value equal to 600 or greater. This signal value is consistently

in a range that is well separated from the central tendency of nontissue voxels in our data. Power spectra were determined (by multiplying the DFT of each time series by its complex conjugate) for each brain voxel. These voxel power spectra were averaged across the brain voxels to determine a "voxel-averaged power spectrum," or spatially averaged power spectrum, per dataset.

Equipment Contributions to fMRI Noise

To provide background data for the interpretation of the results from the human noise datasets, scans of water phantoms were performed in the presence or absence of various pieces of electronic equipment. These experiments were not meant, however, to provide quantitative baseline measures of noise. Two experiments of this type were performed. The purposes of Experiment 1 were (a) to see if computers typically used for stimulus presentation and recording of subject response contribute any noise to fMRI time series and (b) to see if there was any difference between computers in any such noise contribution. The purpose of Experiment 2 was to see if an additional piece of equipment, an overhead projector/LCD panel (InFocus systems panelbook 550) unit, contributed any noise to fMRI time series. Within each experiment, the power spectra were compared across conditions for effects of the equipment on noise.

Experiment 1

Some conditions of the experiment involved manipulating which particular computer (which was located outside of the scanning room) fed a cable to the LCD panel. This cable had a 64-MHz filter installed and was also grounded to the Faraday cage that surrounded the scanning room. Three computers were tested: a Macintosh 280c (portable; this was the computer used during the collection of the primary group of human noise datasets), a Power Macintosh 7100/80 (desktop), and a Macintosh PowerBook 5300c (portable). There was an additional condition in which the cable ran from near (2–4 ft) the Macintosh 280c (turned on) into the LCD panel, but the cable was not actually connected to the computer. The purpose of this condition was to test the idea that the cable could act as an antennae for RF noise, even when not connected to a computer. A control condition was one in which there was no cable running into the room. In all conditions in Experiment 1 (including the control) the overhead projector/LCD panel unit was inside the scanning room and plugged in inside the scanning room. A water phantom was scanned five times under each condition with the same parameters as the human subjects, except that a total of 30 gradient echo–echoplanar images in time were obtained in each of five slices. The data that will be presented for this experiment were processed with the method of motion compensation described above (i.e., the method that was used in the primary group of human noise datasets).

Experiment 2

In one condition the overhead projector/LCD panel unit was inside the scanning room and plugged in there. A second condition did not have the overhead projector/LCD panel unit inside the scanning room. A water phantom was scanned three times under each condition with the same parameters as the human subjects, except that a total of 80 gradient echo–echoplanar images in time were obtained in each of five slices. The data that will be presented for this experiment were processed with the method of motion compensation described above.

Physiological Contributions to fMRI Noise

To determine if the structure (not the magnitude) of noise differed between human subjects and water phantoms, a single human subject (normal male, age 26, who was separate from the primary group of human noise datasets) and a water phantom were scanned on the same day. Specifically, the question of whether nonwhite noise in fMRI time series was specific to humans was addressed. In this experiment, there was no stimulus presentation equipment inside the scanning room for either the phantom (this was one of the conditions from Experiment 2, above) or the human subject. The human subject was scanned three times under each condition with the same parameters as the primary group of human subjects, except that a total of 80 gradient echo–echoplanar images in time were obtained in each of 16 slices. The data that will be presented for this experiment were processed with the method of motion compensation described above.

Effect of Motion Correction on fMRI Noise

To determine the effects of motion correction on the structure and magnitude of fMRI noise, the same human noise dataset described directly above was analyzed under four different conditions of motion correction: (1) no motion correction, (2) the method of motion compensation described above (that was performed on the primary group of human noise datasets, (3) two iterations of least-squares, 6-parameter, rigid-body realignment using SPM96 (Friston *et al.,* 1995a), and (4) condition 3 followed by condition 2. The method used in condition 3 has been shown (Friston *et al.*, 1995a) to be very similar to the frequently employed AIR technique (Woods *et al.,* 1992; Jiang *et al.,* 1995). The purpose of condition 4 was to see if there were any additive effects of conditions 2 and 3. Additive effects would suggest that the two methods were affecting different signal components. For conditions 3 and 4, the

images from all three separate runs (80 images per run) were aligned to the first image of the first run. Correction for spin history (Friston *et al.*, 1996) was not performed.

False-Positive Rate Comparisons

Several versions of the GLM were applied to the noise datasets to empirically determine false-positive rates for an assumed behavioral paradigm effect. The behavioral paradigm effect in these analyses is "assumed" because the subjects were not actually engaged in an experimental paradigm. The temporal structure of the assumed paradigm was designated to be a boxcar with a fundamental frequency of 0.0125 Hz (40 s off/40 s on), which is in a range typical of that used in our and other labs. The different GLM versions varied in terms of temporal smoothing (i.e., temporal autocorrelation introduced in processing), of assumed intrinsic autocorrelation, and of inclusion of a global signal covariate.

Two different, but theoretically related, measures were used to characterize false-positive rates in the spatially unsmoothed noise datasets. One (hereafter referred to as FP1) was the percentage of voxels in each dataset with t values greater than the nominal upper $\alpha = 0.05$ threshold. This is a gross measure of the width of the distributions. Thus, the theoretically expected value of FP1 was 5%. The second false-positive measure (hereafter referred to as FP2) was the proportion of datasets that contained at least one voxel with a t value whose absolute value was greater than the two-tailed nominal $\alpha = 0.05$ Bonferroni corrected threshold. Thus, the theoretically expected value of FP2 was also 5%. FP2 may differ in information from FP1 if the temporal autocorrelation structure is not stationary in space. For example, if there is a relatively small subset of voxels in each dataset that, for whatever reason, has a much greater ratio of [power at the paradigm frequency]:[power at other frequencies] compared to the rest of the dataset, then there will be many datasets that contain voxels that surpass the Bonferroni corrected threshold even though the distribution of test statistics may not be that much broader than theoretically expected. This is because the autocorrelation model was derived from whole-brain voxel-averaged power spectra and thus assumed temporal autocorrelation that was stationary in space. The hypothesis that the central tendency of FP1 was different from 5% was tested with the signed rank test (two-tailed). The hypothesis that FP2 was greater than 5% was tested with the binomial distribution (one-tailed, as there was no power to test that the FP2 was less than 5%). Both hypotheses were tested in each version of the GLM.

Components of the General Linear Model

The modified GLM presented by Worsley and Friston (1995) was used in all of the analyses. In their terminology, $\mathbf{X}^*$ is the dependent variable after all temporal filtering has been applied, $\mathbf{K}$ is a convolution matrix representing all assumed temporal autocorrelations (both intrinsic and exogenously introduced during processing) present in $\mathbf{X}^*$ under the null hypothesis, and $\mathbf{G}^*$ is the matrix of independent variables used in the regression.

It must be clear that we were not testing the theoretical validity of the modified GLM of Worsley and Friston (1995) in these analyses. Rather, we were testing the extent to which the properties of the particular temporal filter used in processing sufficed to describe the net temporal autocorrelation of BOLD fMRI data such that the observed statistical behavior was close to theoretical distributions. We also were testing the effect of the inclusion of estimates of intrinsic autocorrelation on the validity of the model.

A schematic diagram of the time/frequency filtering of the components of the regression model is shown in Fig. 1. Time/frequency filters were used in some analyses to smooth the data. The purpose of the temporal smoothing was to mimic a real analysis in which smoothing with a filter that closely matches the hemodynamic response function maximizes signal to noise (as proposed by Friston *et al.*, 1995c). One type of smoothing kernel used was an estimate of the BOLD IRF obtained from our behavioral dataset as described above. The other was a Poisson kernel (parameter = 8 s) taken from Friston *et al.* (1994). The purpose of using two different filters was to see if the stringency of the low-pass filter impacted the validity of the GLM when intrinsic autocorrelation was ignored (as our IRF estimate passes more high frequencies than the Poisson). The same filters were also applied to the idealized neural response (a boxcar in this case) twice (once using both magnitude and phase information and once using only magnitude information, Fig. 1B) to match the double filtering of any putative fMRI response (once by intrinsic physiology, once in processing). This double filtering is more important for signal detection than for a false-positive analysis, but was done for completeness and comparability to analyses that, at other times, would be performed on behavioral datasets (see companion paper, Aguirre *et al.*, 1997).

A notable addition to the method described in Worsley and Friston (1995) is the use in some analyses of an empirical estimate of the intrinsic autocorrelation structure of fMRI noise data in $\mathbf{K}$. As originally described, the $\mathbf{K}$ matrix contained only the effects of the filtering of the smoothing kernels used in processing. In addition to the effect of smoothing kernel, the effect of intrinsic autocorrelation was modeled in $\mathbf{K}$ in some analyses. Our estimate of intrinsic autocorrelation was obtained from the power spectra of the fMRI noise data. A model

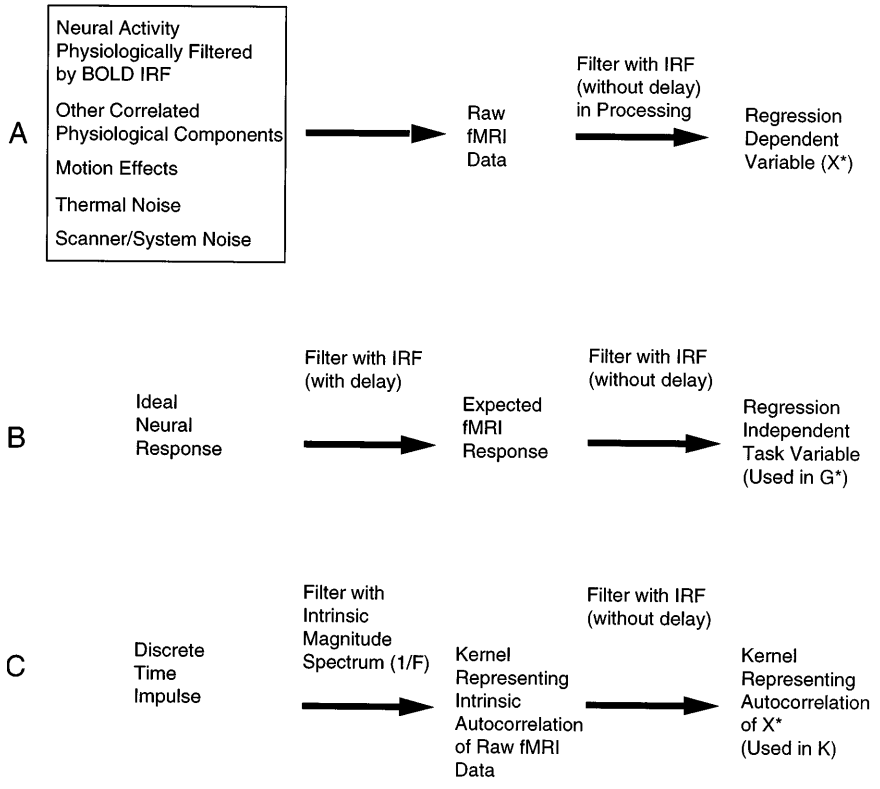

FIG. 1. A schematic diagram of the steps performed to analyze the fMRI data in this report. The statistical framework used was the one presented by Worsley and Friston (1995). (A) Outline of the temporal filtering of the fMRI data. This filtering includes all theoretical contributions from physiology and the scanning system (i.e., intrinsic autocorrelation, shown in the box) and exogenously applied (i.e., applied during processing) steps. (B) Outline of the generation of the behavioral paradigm independent variable used in the regression model. (C) The generation of the **K** matrix, which should reflect all autocorrelation present in the dependent data under the null hypothesis (including both intrinsic and exogenous). The terms "with delay" and "without delay" refer to the use or neglect of, respectively, phase information during the filtering (convolution) process. Phase is neglected during convolution by using the magnitude of each Fourier coefficient of the filter instead of using the complex values. All of these steps, except those shown in the box in (A), were explicitly performed in this report.

of the form

$$|a(\omega)| = \frac{1}{k_1\left(\frac{\omega}{2\pi} + k_2\right)} + k_3, \qquad (1)$$

where $|a(\omega)|$ is the magnitude of the DFT at frequency ω, and k_1 through k_3 are model parameters, was used to fit different noise spectra using iterative, nonlinear least squares ("curvefit" routine IDL, Boulder, CO). These spectra included the square root of the grand average (across the first 13/17 of the noise datasets) of the voxel-averaged power spectra, as well as the square root of the voxel-averaged power spectrum of each individual dataset. This model was chosen because it fit the data best in preliminary datasets. This function of frequency was DFT^{-1} to obtain a time-domain kernel representing the intrinsic autocorrelation of fMRI data. This kernel was incorporated into **K**. If a particular statistical analysis also involved smoothing the data, then the kernel representing intrinsic autocorrelations was itself convolved with the smoothing kernel to obtain the final **K** matrix (Fig. 1C). The purpose of using a subset (13/17) of the total number of noise datasets to generate the autocorrelation model was to allow a qualitative testing of the reliability of this model in the remaining (4/17) subjects.

Separate model parameters for Eq. (1) were determined for analyses that included the global signal as a covariate. The rationale was that the **K** matrix should reflect autocorrelation under the null hypothesis, and that when the model included a global signal covariate, the null hypothesis would be true in data that did not have this signal component. In these cases, the parameters for Eq. (1) were obtained from voxel-averaged power spectra from datasets that had the effects of global signal removed from each voxel via the use of partial correlations.

A decaying exponential model of the form,

$$|a(\omega)| = k_1 \cdot e^{-\omega/k_2 \cdot 2\pi} + k_3, \qquad (2)$$

was also fit to the voxel-averaged power spectrum (and its square root) of each individual dataset. This model was chosen to compete with Eq. (1) as it is also a decreasing function with a positive second derivative

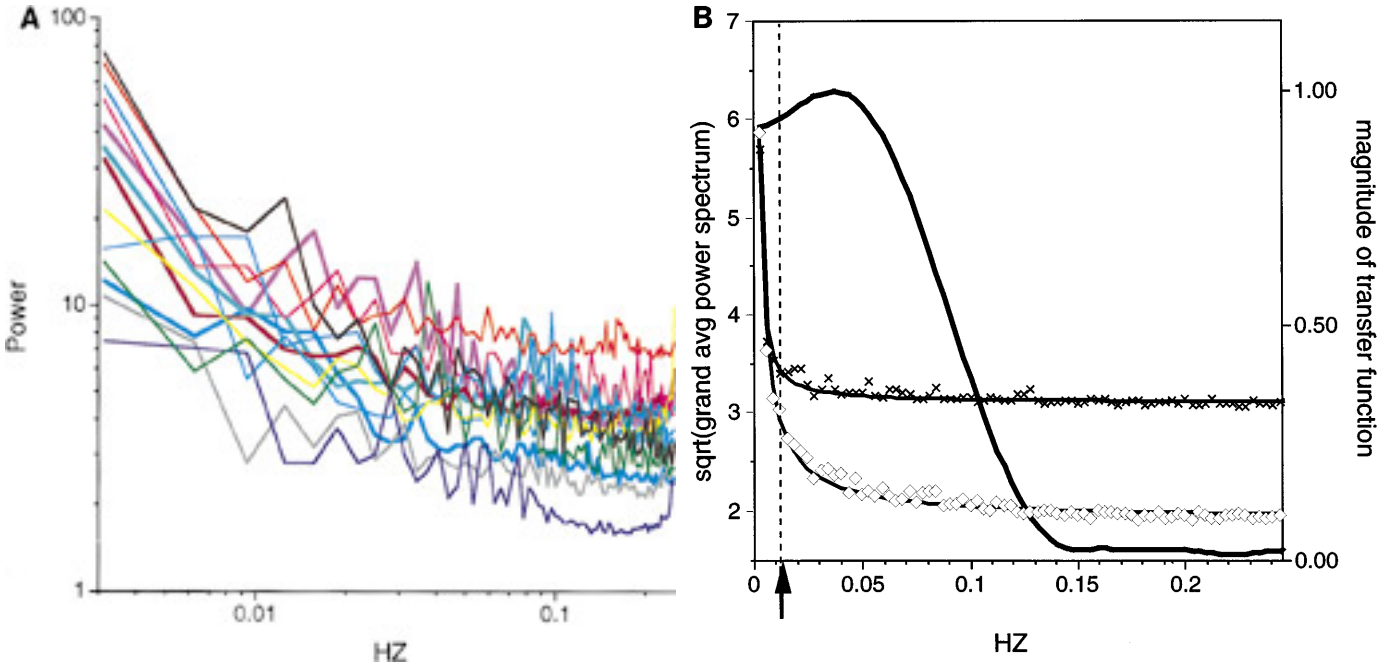

FIG. 2. Power spectra of human BOLD fMRI "noise" data. (A) plot of the voxel-averaged power spectra vs temporal frequency for 13/17 of the noise datasets. The power spectrum of each dataset is shown in a different color. A general pattern of increasing power with decreasing frequency is consistent, but the degree of this relationship seems to vary between datasets. The x and y axes are $\log_{10}$ scaled. (B) A plot of the square root of the grand average power spectrum vs temporal frequency for 13/17 of the noise datasets is shown (open diamonds) with its fit to Eq. (1). In addition, the analogous data for the remaining 4/17 noise datasets is shown (small crosses) with its fit. The magnitude of the frequency representation of our empirically derived impulse response function is also shown for comparison (thick solid line). The arrow and vertical dashed line denote the fundamental frequency of the boxcar paradigm assumed in the statistical analyses (0.0125 Hz).

and horizontal asymptote. The explanatory power of (2) was compared to that of (1) by comparing the difference in mean square residuals for the fits of each dataset. These differences (17 total) were tested against 0 with a sign test (two-tailed, $\alpha = 0.05$). If one model consistently fit better than the other, then it would tend to have a lower residual error.

Discrete time sines and cosines periodic with the length of the time series (160 images) and with frequencies below that of the fundamental frequency of the assumed paradigm were included as covariates in all of the analyses. Their shape does not change with linear filtering as they are eigenmodes of linear systems. Their inclusion high-pass filters the data to remove poorly understood low-frequency components of the data (Friston *et al.*, 1995b). Even though we have attempted to model power in all frequencies with the empirically estimated autocorrelation, power in the lowest frequencies seems to be more variable than power in the higher frequencies (this can be observed in Fig. 2A), and it was decided to remove their effects from the analysis in this way. The global signal (obtained as explained above) was included as a covariate in some analyses.

Some analyses involved a "straight" boxcar correlation with no assumed autocorrelation (i.e., assumed independence), but were still performed using the structure of the GLM of Worsley and Friston (1995) to maintain rigid experimental control over any differences between the analyses. This was easily accomplished as the model can accommodate any assumed autocorrelation structure by using the appropriate **K**. In the assumed independence case, the kernel of **K** was a discrete time impulse. The purpose of performing the boxcar correlation was to compare the results of this standard method of fMRI analysis (Bandettini *et al.*, 1993) with methods that take autocorrelation into account.

Estimation of Smoothness Contributions Representable by a Continuously Differentiable Spatial Autocovariance Function

The statistical maps generated from each human subject noise dataset (under the analysis conditions of $1/f$ model included in **K**, data smoothed with our estimated IRF, and global covariate not in the model) were subjected to in-plane (i.e., 2-D) smoothness estimation using the method of Xiong *et al.* (1995). Their measure of smoothness was $|\Lambda_t|$,

$$|\Lambda_t| = \left[\frac{2\sqrt{|\Lambda_e|}}{2 + \sqrt{|\Lambda_e|}}\right]^2, \qquad (3)$$

where $|\Lambda_e|$ is the determinant of Λ_e, the estimated variance–covariance matrix of the partial derivatives of the slice (Xiong *et al.*, 1995). Methods that assume that the smoothness of the data is large compared to voxel size (Friston *et al.*, 1991; Worsley *et al.*, 1992) use $|\Lambda_e|$ as the measure of smoothness instead of $|\Lambda_t|$. $|\Lambda_t|$ was determined in each slice of each dataset (using map-wise variances, see Aguirre *et al.*, 1997), and an average in-plane $|\Lambda_t|$ per dataset was generated. The $|\Lambda_t|$ was adjusted to be appropriate for describing the underlying component processes of the statistical maps (Worsley *et al.*, 1992; Holmes, 1994). Each of these $|\Lambda_t|$ was then converted into an effective FWHM for the corresponding Gaussian kernel that would have led to that $|\Lambda_t|$ if used to convolve a pixelated slice of multivariate Gaussian data (see Appendix of Xiong *et al.*, 1995):

$$\mathrm{FWHM} = [4 \ln (2)(|\Lambda_t|^{-1/2} - 1)]^{1/2}. \quad (4)$$

This FWHM measure is not affected by pixelation and will be nearly 0 for a spatially uncorrelated dataset.

Global Signal and Spatial Coherency Measurement

The global signal for each dataset was determined by computing the average time series across all brain voxels. Power spectra were determined for these global signals. If expected cross-correlations between voxels are 0 (i.e., if off-diagonal elements of the voxel variance–covariance matrix are expected to equal 0), then the power spectrum of the global signal is expected to equal the voxel-averaged power spectrum divided by the number of voxels. If expected cross-correlations between voxels are unity (i.e., all off-diagonal elements of the voxel variance–covariance matrix expected to equal 1), then the power spectrum of the global signal is expected to equal the voxel-averaged power spectrum. Based on this idea, a measure of spatial coherency (SC) vs temporal frequency was determined for each dataset,

$$\mathrm{SC}(\omega) = \frac{\left[\dfrac{\mathrm{PS}_{\text{global signal}}(\omega) \cdot V}{\mathrm{PS}_{\text{voxel averaged}}(\omega)}\right] - 1}{V - 1}, \quad (5)$$

where $\mathrm{PS}_{\text{global signal}}(\omega)$ is the power spectrum of the global signal at temporal frequency ω, $\mathrm{PS}_{\text{voxel averaged}}(\omega)$ is the voxel-averaged power spectrum at temporal frequency ω, and V is the voxel count for the dataset. $\mathrm{SC}(\omega)$ was expected to be 0 if the expected cross-correlation at zero-lag between voxels at temporal frequency ω was 0. $\mathrm{SC}(\omega)$ was expected to be 1 if the expected cross-correlation at zero-lag between voxels at temporal frequency ω was 1. Intermediate values would be expected under intermediate amounts of expected cross-correlation.

It is worthwhile to consider the differences between the SC measure of Eq. (5) and the FWHM measure of Eq. (4). One difference between FWHM and SC is that FWHM (as implemented here) examines spatial smoothness primarily at one temporal frequency, the fundamental frequency of the behavioral paradigm frequency used in generating the statistical map. In contrast, SC is a measure of spatial smoothness vs temporal frequency. It may come as a surprise to some that spatial smoothness can vary as a function of temporal frequency. However, if one considers each voxel time series as a composition of sines and cosines at different temporal frequencies, one can imagine that across voxels the sines and cosines at some frequencies may be more in phase than at other frequencies. The second difference between FWHM and SC is that FWHM measures components of spatial smoothness that can be represented by a continuously differentiable autocovariance function. It is this component that is generally referred to as "smoothness" when discussing map-wise statistical methods like SPM (Friston *et al.*, 1991; Worsley *et al.*, 1992). Though SC is also sensitive to this component of smoothness (under most circumstances, e.g., if one smooths with a Gaussian kernel), it is also sensitive to other components of smoothness, namely those that cannot be represented by a continuously differentiable autocovariance function. An example of a situation in which smoothness could not be represented by a continuously differentiable autocovariance function would be in a dataset which comprised a spatially uncorrelated component plus a signal component shared by all voxels. In this case all the off-diagonal elements of the voxel variance–covariance matrix would be expected to have the same positive value. The FWHM of such a dataset would be 0. However, SC would detect this spatial coherence. Thus, the use of both FWHM and SC allowed the examination of different components of spatial smoothness at the paradigm frequency.

The global signals were correlated voxel-wise with their datasets of origin for several subjects to create correlation maps. These maps represent the spatial distribution of the relative presence of the global signals throughout the volumetric datasets. These maps were thresholded at an arbitrary $|R| > 0.2$ and were used to qualitatively test the idea that the global signal segregated strongly to particular parts of the dataset (i.e., gray matter, white matter, or ventricle) or strongly resembled motion artifact (in which case we would expect it to segregate to regions with large signal gradients).

RESULTS

Characterization of Intrinsic Temporal Autocorrelation

Both the $1/f$ model [Eq. (1)] and the decaying exponential model [Eq. (2)] converged for the voxel-averaged

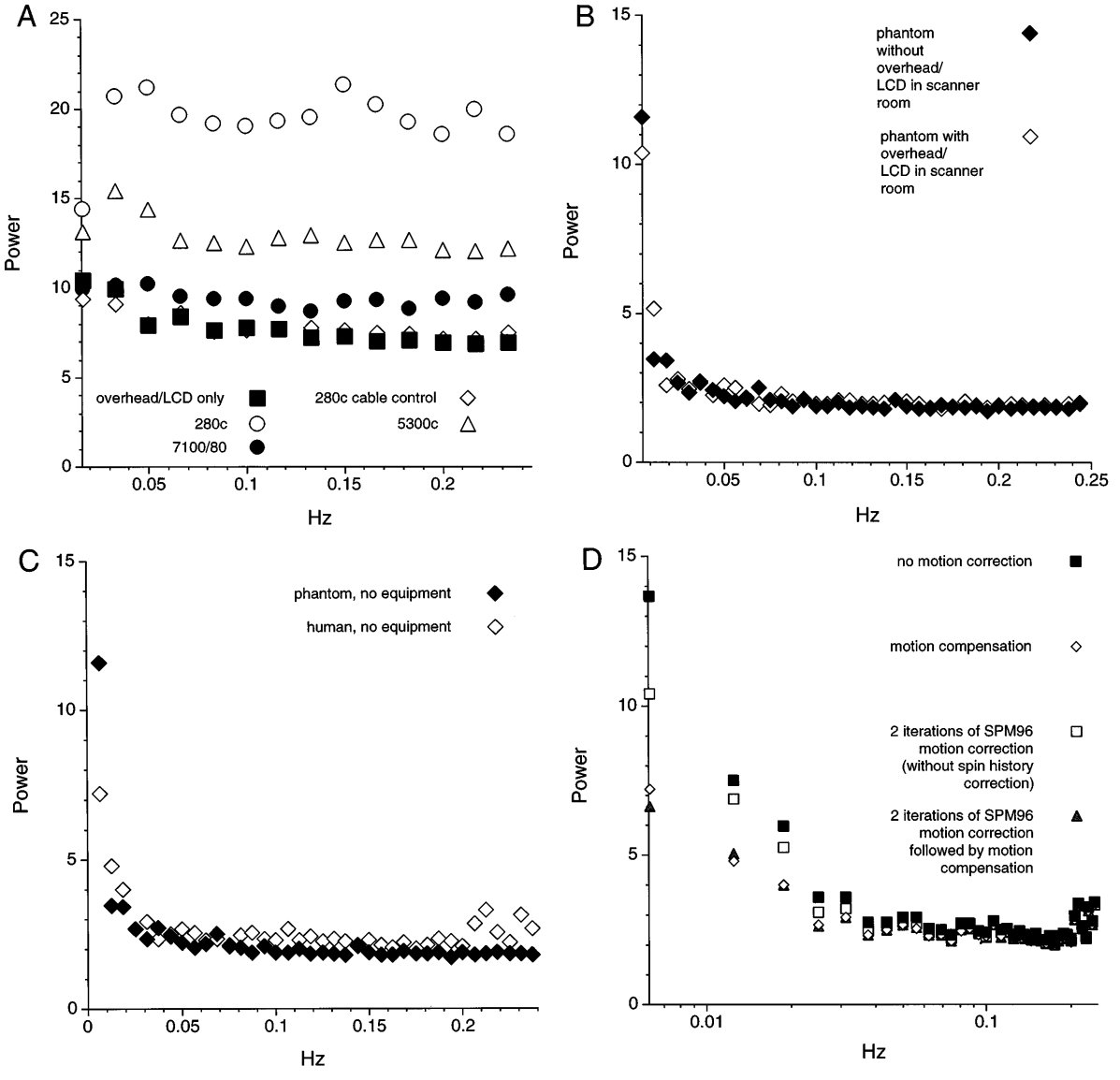

FIG. 3. (A) The effect of various stimulus presentation computers on noise in a water phantom is shown. Different computers contributed different amounts of white noise. (B) The effect of the overhead projector/LCD unit on noise in a water phantom is shown. No effect was apparent. (C) The structures of the noise in a human and in a phantom are shown. Both human and phantom evidenced a $1/f$ component. (D) The effect of motion correction procedure on the shape and magnitude of the power spectrum of human "noise" is shown (no stimulus presentation equipment was present during the acquisition of these data). Motion correction seems to preferentially reduce noise at lower frequencies. Also, our method of motion compensation seemed to affect a superset of the variance components reduced by the rigid-body transformation. The x axis is $\log_{10}$ scaled in D to better display the lower frequency range. All power spectra are averaged across space and datasets ($n = 3$ datasets per condition).

power spectra of each of the 17 human noise datasets. However, Eq. (1) tended to model more of the variance than Eq. (2) [15/17 mean squared residuals were less for Eq. (1) than for Eq. (2), $P < 0.0023$, Sign test). This result, in conjunction with the absolute adequacy visually of the $1/f$ model (Fig. 2B), led us to accept a $1/f$ model over a decaying exponential model as the form of the frequency domain description of the intrinsic autocorrelation of BOLD fMRI data.

The voxel-averaged power spectra of 13/17 noise datasets are shown in Fig. 2A. Even though Eq. (1) fit the voxel-averaged power spectra of all of the datasets reasonably, all the voxel-averaged power spectra did not appear to be realizations of the same process. No statistical test was performed, however, to rigorously test this idea. In addition to the $1/f$ component, there was a white noise component to the signal as seen in Fig. 2A (also observed by Weisskoff *et al.*, 1993, and Boynton *et al.*, 1996). This is expected as the fMRI signal is not purely physiological in origin but also has

contributions from other noise sources. The square root of the grand average of these same power spectra is shown in Fig. 2B, along with the nonlinear, least-squares fit using Eq. (1). This particular fit was used as a frequency domain representation estimate of the expected intrinsic autocorrelation of human fMRI data for our system (when the global signal was not used as a covariate, see Methods).

The square root of the grand average of the other 4/17 noise dataset power spectra and its fit are also shown in Fig. 2B. It can be seen that though a $1/f$ model of the form given in Eq. (1) still fits these data adequately, the shape is different from the fit obtained with the other 13/17 datasets. It is apparent that the 4/17 datasets had a greater white noise component than the 13/17, on average. There was indirect evidence that this discrepancy was due to the appearance of an RF noise source from the stimulus presentation computer (see Methods) toward the end of the several months taken to collect these datasets. We now present the results of studies aimed at characterizing contributions to our observed noise spectra from equipment, physiology, and motion correction procedures, respectively.

Characterization of Contributions to Noise Spectra

The primary group of 17 human noise datasets was acquired in the presence of stimulus presentation equipment to mimic conditions that would exist during the performance of behavioral paradigms. This equipment included an overhead projector/LCD panel unit and a Macintosh 280c computer. As mentioned above, the white noise component present in the last datasets collected was higher than in earlier datasets, suggesting the emergence of a new RF noise source. Though this idea could not be tested directly, a study using a water phantom showed that the use of computers in general during scanning did increase white noise (this study was performed after the collection of the 17 human noise datasets; Fig. 3A). In addition, the magnitude of the noise depended upon the specific computer used, with the computer used in the primary group of human noise studies (the 280c) showing the greatest noise of the three tested. A control condition in which a cable that fed into the scanner room was not hooked up to any computer showed that the cable was not acting as a nonspecific antennae for RF noise. In contrast to the computer contributions, the overhead projector/LCD panel unit did not increase noise (Fig. 3B).

We wished to determine if the $1/f$ component of the noise observed in the human subjects was necessarily due to a physiological cause. An average (across three runs) human noise power spectrum (from a single subject, separate from the primary group of 17, and acquired in the absence of any equipment) was compared to that of a phantom (the same one whose data are presented in Figs. 3A and 3B). Both evidenced a $1/f$ component (Fig. 3C). This shows that the presence of a $1/f$ component does not imply a physiological source.

The results presented above (in Figs. 2A, 2B, and 3A–3C) are for data processed with the motion compensation method presented under Methods. This procedure is not a true motion correction (as explained under Methods). We wished to determine if the presence of the $1/f$ component could be attributable completely or in part to motion due to the use of a possibly ineffectual motion correction method. Figure 3D compares voxel-averaged power spectra for a single set of human "noise" data (the same presented in Fig. 3C) processed with a true motion correction method (the least-squares, 6-parameter, rigid-body realignment routine of SPM96) to that processed with our motion compensation method. The power spectrum resulting when no motion correction was applied is also shown, as well as that resulting from SPM96 motion correction followed by our motion compensation method. This last condition was performed to see if there were any additive effects of the two treatments. The effect of motion correction/compensation on the power spectra was larger at the lower temporal frequencies. Data processed in the absence of motion correction yielded the largest $1/f$ component (relative to the flat frequency response), suggesting that motion makes a contribution to this component. However, data processed under each of the motion correction procedures, including the combined method, evidenced a $1/f$ component. The lack of a substantial additive effect of SPM96 motion correction and our method suggests that they affect similar signal components and that, by inference, the $1/f$ component observed in the primary group of 17 subjects (processed with our motion compensation method alone) could not be attributed to the first-order effects of motion.

Comparison of BOLD IRF Estimate with Intrinsic Autocorrelation Structure

An estimate of the neural activity to BOLD IRF obtained from the behavioral dataset is shown in time in Fig. 4. Neither the filtering characteristics of this IRF estimate nor those of other putative IRFs obtained from activation data reported (Bandettini, 1993, as presented in Friston *et al.*, 1994; Boynton *et al.*, 1996) are sufficient to explain the observed intrinsic autocorrelation in BOLD data. There are two remarkable differences between the IRFs and the observed intrinsic autocorrelation. One such difference is the presence of a white noise, or flat frequency response, component. This is not unexpected as only those variance components of the fMRI time series that conformed to convolution of the underlying neuronal processes would have temporal autocorrelations corresponding to the IRF (Friston *et al.*, 1994). It is expected that there are uncorrelated components to the fMRI signal, such as

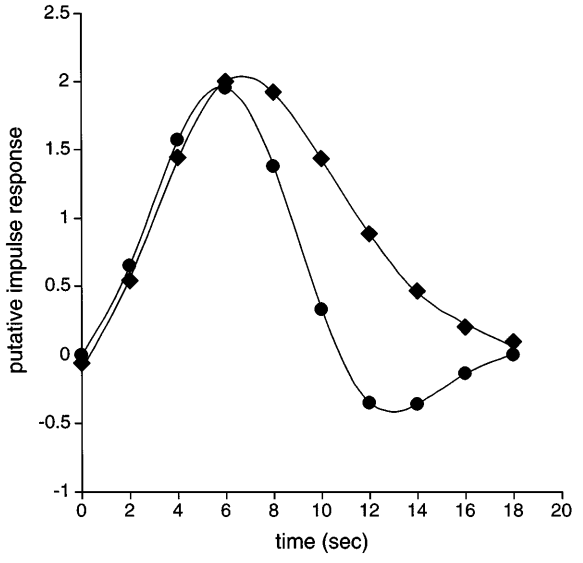

FIG. 4. Two putative BOLD impulse response functions are shown. Our estimated IRF, obtained from a behavioral dataset, is represented by black circles. The IRF proposed by Friston *et al.* (1994), modeled as a Poisson function with a parameter of 8 s (that has been deconvolved with a Gaussian, $\sigma = 1.9$ s to account for processing performed in Friston *et al.,* 1994), is represented by diamonds. The Poisson IRF estimate is broader in time and thus cuts off at a lower frequency than our IRF estimate.

scanner noise, in addition to the autocorrelated components evoked by neural processes.

A second, perhaps more unexpected, difference between the IRFs and the observed intrinsic autocorrelation is that the IRF filtering nearly levels off in the range of frequencies typically used as task frequencies (0.0083–0.025 Hz) while the observed intrinsic autocorrelation in the noise is sharply rising in this range (Fig. 2B). It thus appears that the 1/*f* component we have observed in human "noise" data cannot be explained by the convolution of neuronal processes (Friston *et al.*, 1994) by a low-pass filter akin to those reported in the literature from activation data (Bandettini, 1993; Boynton *et al.*, 1996).

The Poisson kernel (parameter = 8 s) suggested by Friston and colleagues (1994) is also shown in Fig. 4 for comparison with the IRF estimate obtained here. The Poisson is slightly broader in time and is thus a more stringent low-pass filter than the BOLD IRF estimate. An explanation of why these two kernels are different is that they were obtained with different methods that relied on different assumptions. The method used here to obtain an IRF estimate (that is, using fMRI activation responses) made two assumptions, (1) that the system mediating the transformation of neural signal to fMRI signal change was linear-time invariant and (2) the change in neural input was approximately a step function. The first assumption has recently been shown not to hold for short durations of stimulation though it may be a reasonable approximation for longer durations of stimulation (Vasquez and Noll, 1996; Boynton *et al.*, 1996). The second assumption was not tested, but we feel confident that neural dynamics do respond faster than hemodynamics, thus ensuring that the initial rise time seen in the BOLD IRF estimate is not due to a neural response blurred on the time scale of seconds. In contrast, Friston and colleagues (1994) used the phase-shifted differences of fMRI time series to estimate the IRF. In doing so, Friston and colleagues made three assumptions: (1) Again, that the system mediating the transformation of neural signal to fMRI signal change was linear-time invariant, (2) that the intrinsic autocorrelation present in fMRI data was due entirely to the transfer function of said linear system, and (3) that the transfer function would be modeled well by a Poisson in time. It would seem that assumption 2 is not supported by the data presented here. Other data showing the disparity between the functional response of the BOLD fMRI signal and the temporal autocorrelation present in the fMRI data have been recently reported by Boynton and colleagues (1996).

The Effect of Assumed Temporal Autocorrelation on False-Positive Rates

The fMRI noise datasets were analyzed with versions of the GLM that varied in their assumed temporal autocorrelation (that was modeled in the **K** matrix). False-positive measures from these different analyses were compared to theoretically expected (i.e., tabular) false-positive rates. In all the analyses described in this section, the global signal was included as a covariate because (1) the purpose of the analyses of this section was not to determine the effect of the global signal (thus obviating the need to describe all of the analyses with and without the global signal, as there were no apparent interactions) and (2) its inclusion as a covariate was judged to increase validity in other analyses that did examine its effect (presented in the following section).

When independence was assumed, both false-positive measures were significantly greater than theoretically expected (median FP1 = 6.4%, $P < 0.023$; FP2 = 10/17 maps, $P < 1.4 \times 10^{-9}$). This type of regression analysis corresponds closely (except for the inclusion of the global signal covariate and low frequency confounds) to the "straight" boxcar correlation. These results suggest that the use of this type of analysis should not be expected to yield false-positive rates in accord with tabular values. The discrepancy was particularly egregious when considering map maxima (i.e., FP2).

If the 1/*f* autocorrelation model obtained from the square root of the grand average power spectrum correctly describes the expected intrinsic autocorrelation, then its inclusion in the **K** matrix should bring the

false-positive rate close to the tabular value. This statement was confirmed with simulations on computer-generated, uncorrelated noise that was filtered with a particular $1/f$ model and then analyzed with the modified GLM using this same kernel in the **K** matrix (data not shown). However, when the empirically estimated autocorrelations were used as the model on the actual noise data, a paradoxical effect was observed. FP1 was significantly less than tabular values, while FP2 was significantly greater than tabular values (median FP1 = 3.1%, $P < 0.0023$; FP2 = 5/17 maps, $P < 0.0012$). This result supports the use of both false-positive measures, and calls into question the validity of the autocorrelation model. While the autocorrelation model fits the average power spectra quite well (Fig. 3B), the average was across both voxels and datasets. The model would only be valid if the expected autocorrelation was the same in every voxel in every dataset. As FP1 is significantly less than tabular values, the autocorrelation model is systematically overestimating intrinsic autocorrelation on a voxel-wise basis. However, that FP2 is greater than tabular values shows that the autocorrelation model is systematically underestimating intrinsic autocorrelation in at least some subpopulation of voxels in at least some of the datasets. The use of voxel-averaged power spectra to derive a model of intrinsic autocorrelation (the method used here) assumed that the expected temporal autocorrelation was stationary in space (i.e., every voxel had the same expected temporal autocorrelation). This result shows that this initial assumption was not satisfied. However, the inclusion of the estimate of intrinsic autocorrelation in the analysis did reduce both FP1 and FP2 compared to the assumed independence case, as expected.

When the data were smoothed with our empirically derived BOLD IRF, and the $1/f$ model of intrinsic autocorrelation was included, FP1 was significantly less than tabular values while FP2 was not greater than tabular values (median FP1 = 3.1%, $P < 0.013$; FP2 = 1/17 maps, $P < 0.58$). In an otherwise identical analysis, which ignored intrinsic autocorrelation, FP1 was significantly greater than tabular values, while FP2 had a trend toward being greater (median FP1 = 5.9%, $P < 0.013$; FP2 = 3/17 maps, $P < 0.0503$). Thus, when smoothing with this filter and ignoring intrinsic autocorrelation, FP1 was significantly greater than theoretically expected, while FP2 had a trend in the same direction. When the intrinsic autocorrelation was taken into account, the false-positive rates were brought to a level that was not greater than tabular values, though FP1 was now significantly less than tabular values.

The final two analysis versions employed a different smoothing kernel: a Poisson kernel with parameter = 8 s (Friston *et al.*, 1994). When this kernel was used and intrinsic autocorrelation was taken into account, FP1 was significantly less than tabular values (median FP1 = 1.8%, $P < 0.0003$; FP2 = 0/17 maps, $P < 1$). When intrinsic autocorrelation was ignored, neither FP1 nor FP2 differed from expected (median FP1 = 5.4%, $P < 0.332$; FP2 = 2/17 maps, $P < 0.208$). Thus, filtering with the Poisson kernel yielded false-positive rates at levels not significantly different from tabular values. However, attempting to additionally model intrinsic autocorrelation caused the false-positive rates to fall below tabular values.

The variation in FP1 between datasets was much greater than would be expected if the test-statistic values from all of the datasets were sampled from the same population. If they were from the same population, then we may look at the FP1 of each dataset as a sample proportion and the sample of these measures as a sample of sample proportions. From the binomial distribution, the variance of a sample proportion with true proportion p_o, is $p_o^* (1 - p_o)/n$, where n is the sample size used in calculating each proportion (here equal to the number of voxels per dataset). The smallest n for the datasets (that would thus yield the maximum calculated variance) was about 10,000 voxels. This gives an expected variance of 3.5×10^{-6} when $p_o = 0.036$ (= the mean value of FP1 obtained for the analysis with our empirically derived IRF as the temporal smoothing kernel, intrinsic autocorrelations modeled, and global signal included as covariate), while the sample variance of FP1 (for this particular analysis) was 2.4×10^{-4} (approximately 70 times larger than expected). As this analysis included a global signal covariate, the statistical effect of the global signal could not be the cause of this observation (see below for a description of the statistical effect of the global signal). This suggests systematic variations in statistical properties between datasets, perhaps due to systematic variations in temporal autocorrelation as suggested by Fig. 2A. Another possible explanation is that there were differences across datasets in the extent of the violation of normality.

Global Signal Characterization

A global signal from a randomly chosen noise dataset is shown in Fig. 5A. There is a great deal of structure evident, indicative of temporal autocorrelation. It can be seen that in this particular subject there exists a pattern in time that anti-correlates ($R = -0.33$) with the assumed behavioral paradigm (which is plotted as a reference). The assumed paradigm is an experimental artifice as it did not occur, but its structure in time is typical of that commonly used in fMRI studies. This global signal also provides an example of the relatively high spurious correlations that can be observed ($P < 0.00001$ if independence is assumed) when temporal autocorrelation is present, but not considered.

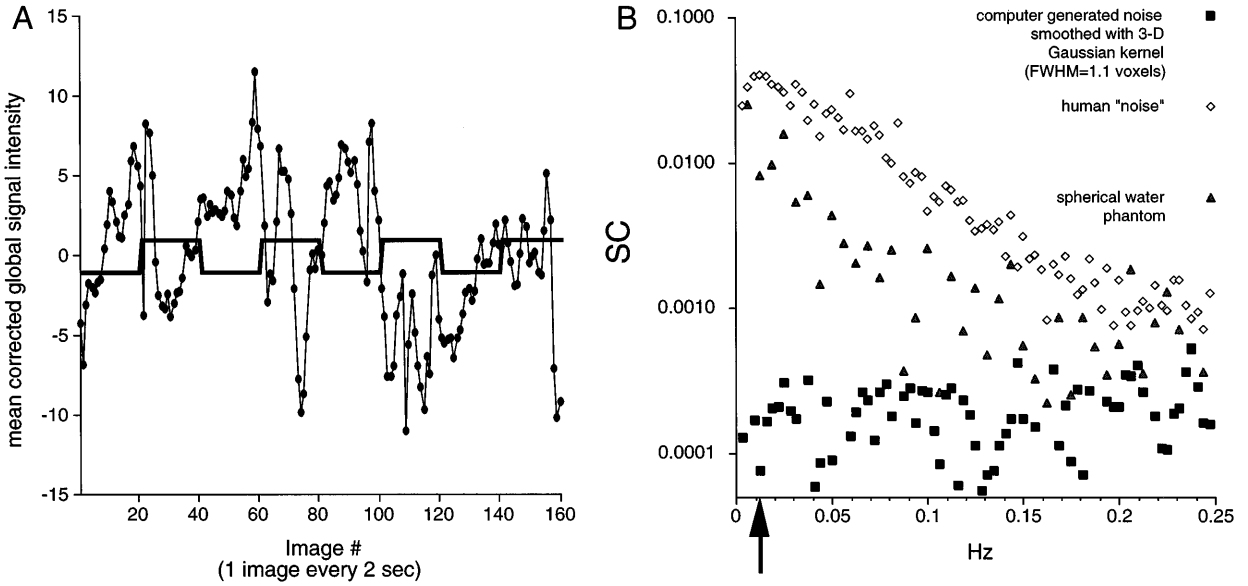

FIG. 5. Characterization of the global signal. (A) Plot of the global signal from a randomly selected noise dataset (circles). Overlaid is a plot of the "assumed" behavioral paradigm (0.0125-Hz square wave). The temporal structure (i.e., autocorrelation) present in the global signal is evident. (B) Shown is the SC measure (defined under Methods) vs frequency. Data for both the human noise datasets (averaged across 17 subjects) and the computer-generated noise (averaged across five $30 \times 30 \times 10$ datasets which were smoothed with a 3-D Gaussian kernel of FWHM = 1.1 voxels) are presented. The presence of SC values >0 at all temporal frequencies in the human noise demonstrates that spatially coherent signal exists in human subject fMRI datasets. That the SC for the human noise datasets at the paradigm frequency (arrow) is two orders of magnitude greater than the SC for computer-generated noise spatially smoothed with a Gaussian kernel of FWHM = 1.1 voxels (which is greater than that of the component fields estimated from the statistical maps of human noise) shows that the SC cannot be entirely explained by a continuously differentiable autocorrelation function. The spatial coherence of human noise is greater at lower frequencies. The *y* axis of B is log10 scaled.

If the expected covariances of all voxel time series in a given dataset are 0 (i.e., no expected cross-correlations between voxel time series at zero lag), then the expected power spectrum of the global signal should be the expected voxel-average power spectrum divided by the number of voxels in that dataset. A measure based on this ratio was used to determine spatial coherency as a function of temporal frequency. The expected value of this SC measure under the null hypothesis was 0 at each frequency, while perfect spatial coherency would yield a value of 1. The SC measure (averaged across the 17 human noise datasets) is shown in Fig. 5B. The ratio is greater than 0 everywhere, showing spatial coherence at all frequencies. The SC measure is greater at low frequencies compared to high, being fit well for most of its range by a decaying exponential. It should be noted that the SC measure in data collected from a spherical water phantom was similar in magnitude and shape to that of the human datasets (Fig. 5B), suggesting that this phenomenon is not completely physiological (assuming that the SC observed in the phantoms is also contributing to the SC of human datasets). Also shown in Fig. 5B is the SC measure of computer-generated noise datasets that were smoothed with a FWHM that was greater than that of the component fields estimated from the statistical maps of human noise (see below). The purpose of performing this simulation was to provide an estimate of the SC measure that would result from spatial smoothness that could be represented as a continuously differentiable autocorrelation function that was greater than or equal to that measured in human noise (see below for these estimates). Smoothing the computer-generated noise with a 3-D Gaussian kernel of FWHM = 1.1 voxels (approximately 2 SDs above the 2-D FWHM of human noise at the paradigm frequency; see below) yielded an SC measure that was approximately two orders of magnitude less than the SC of the human noise near the paradigm frequency. The apparent discrepancy of using an in-plane measure of the inherent smoothness of the human noise datasets and of using a 3-D smoothing kernel on the computer-generated noise was not expected to be important as it was assumed that the smoothness in *z* (in units of voxel size in *z*) would have been if anything less than the in-plane smoothness (due to the greater physical size of the voxels in *z*). The lack of any strong trends in frequency of the SC measure in the computer-generated noise was expected as the spatial smoothing should not have preferentially affected particular temporal frequencies.

An advantage of the measure of spatial coherence described above is its sensitivity to all forms of spatial smoothness. This is also a disadvantage in that different types of smoothnesses cannot be distinguished. One

type of spatial smoothness can be represented by a stationary, continuously differentiable autocorrelation function. This type of spatial smoothness can be increased in a dataset by convolving with a continuously differentiable spatial kernel (e.g., a Gaussian kernel). This is the type of spatial smoothness required by certain map-wise statistical methods, like SPM (Friston *et al.*, 1991; Worsley *et al.*, 1992). The FWHM of the 2-D Gaussian filter that would have yielded the observed smoothness (of the component processes) of this type that was observed in the statistical maps of the noise datasets was determined (Xiong *et al.*, 1995). The mean in-plane FWHM ($\pm$SD) across the primary group of human noise datasets ($n = 17$) was 0.90 ± 0.08 pixels when the analysis did not include a global signal covariate (and 0.88 ± 0.09 pixels when the analysis did include a global signal covariate; see Aguirre *et al.*, 1997, for a discussion of the effect of a global signal covariate on smoothness estimation). This observed FWHM is a rather low value compared to what would be required to yield the SC values that were observed in human fMRI datasets (see Fig. 5B and above). The in-plane FWHM of a spherical water phantom ($n = 3$) was 0.90 ± 0.11 pixels.

To examine the possibility that the global signal is neuroanatomically localized, some of the noise datasets were voxel-wise correlated against their global signals. Three randomly selected correlation maps thresholded at an arbitrary $|R| > 0.2$ are shown in Fig. 6. It is evident that (1) the regions that correlate with the global signal do not seem to segregate strongly to one type of neuroanatomic tissue (i.e., subcortical vs cortical, gray vs white, though white matter seems to be somewhat avoided), though there do seem to be "hotspots" in which there are focal sites of high correlation; (2) positively correlated regions are present throughout the extent of the brain, while there is a paucity of negatively correlated regions (supporting the earlier observation of general spatial coherence); and (3) the correlated regions do not appear similar in appearance to motion artifact (which one would expect to appear along high spatial signal gradients). This last inference is supported by an analysis of the effect of motion correction on the global signal. The global signal from the same dataset as that shown in Fig. 5A was similar (as assessed by raw correlations) when it was processed with different motion correction procedures: $R_{\text{spm96,no motion correction}} = 0.70$, $R_{\text{motion compensation, no motion correction}} = 0.98$, and $R_{\text{spm96,motion compensation}} = 0.82$.

Effect of Global Signal on Statistical Distributions of Spatially Unsmoothed Data

The SC measures presented above showed that the fMRI global signal is a spatially coherent process, not simply the average of uncorrelated voxel time series. Figure 7 shows the voxel-wise distribution of *t* statistics from the same dataset as in Fig. 5A when analyzed with a regression model (with assumed independence, i.e., no modeling of intrinsic autocorrelation) which (1) did not include the global signal as a covariate (black line) and (2) did include the global signal as a covariate (gray line). It can be seen that the distribution is shifted to the left of 0 when the global signal is excluded from the model and that the inclusion of the global signal in the model more or less centers the distribution. The shift to the left of 0 when the global signal was not included as a covariate is due to the negative correlation of the global signal with the assumed paradigm (shown in Fig. 5A). The variation of the mean *t* values of the noise datasets was much greater when a global signal covariate was not included than when it was included [$F(16, 16) = 109$, $P < 2.9 \times 10^{-13}$, descriptive as the samples were not independent]. In addition, including the global signal as a covariate stabilized the number of voxels that exceeded the theoretical $\alpha = 0.05$ threshold [i.e., reduced the variance of the FP1 measure $F(16, 16) = 3.1$, $P < 0.015$, descriptive as the samples were not independent]. Thus, inclusion of a global signal covariate in analysis improved statistical validity by reducing the variation of voxel-wise false-positive rates across datasets. Interestingly, the inclusion of the global signal did not have an effect on the median of FP1, or on FP2 in any of the analysis versions. The lack of effect on FP1 is expected as the global signal's effects should not have been systematic across datasets (as there is no reason to expect a systematic phase relationship between the global signal and the assumed paradigm across datasets), and only systematic effects would have been detected in the central tendency analysis of FP1. The lack of effect on FP2 is not expected *a priori* as it is conceivable that shifts of the distributions could have caused an increase in the number of datasets that exceeded the Bonferroni corrected threshold.

DISCUSSION

It was observed that BOLD fMRI data from both human subjects and phantoms were autocorrelated in time. In particular, lower frequencies (<0.05 Hz) tended to have greater power than higher frequencies. This autocorrelation structure could not be explained by convolution of neural activity (Friston *et al.*, 1994). In addition to such a mechanism, other possible low-frequency contributors to the observed fMRI signal in humans include motion, shifts in B_0 (Jezzard, 1996), respiratory cycles, aliased cardiac cycles, and other, as yet poorly understood, physiological components (Biswal *et al.*, 1996a). However, we have shown that at least the first-order effects of motion cannot completely explain the $1/f$ component. Also, respiratory cycles are not low enough in frequency to account for the pattern ob-

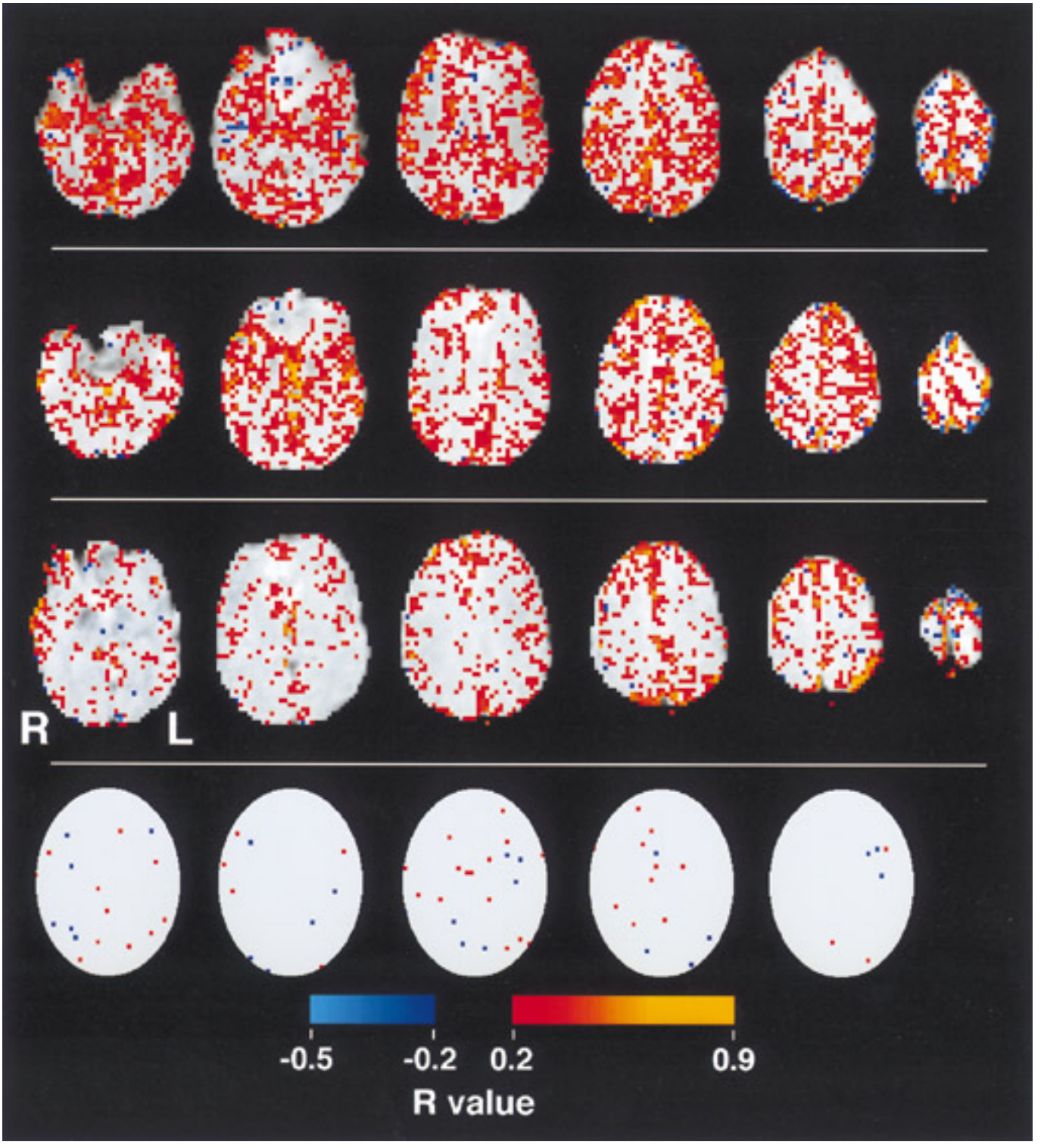

FIG. 6. Shown are global signal correlation maps for three randomly selected human noise datasets (each row shows every third slice from a particular dataset). Shown for comparison on the bottom row is a global signal correlation map for computer-generated noise that has been smoothed with a Gaussian kernel of FWHM = 1.1 voxels. The color map shown at the bottom denotes the strength of the correlation of each voxel in each dataset with its corresponding global signal. The maps are arbitrarily thresholded at $|R| > 0.2$. The spatial distribution of the relative presence of the global signal can be seen from these maps. Note the relative preponderance of positive over negative correlations in the maps derived from human datasets.

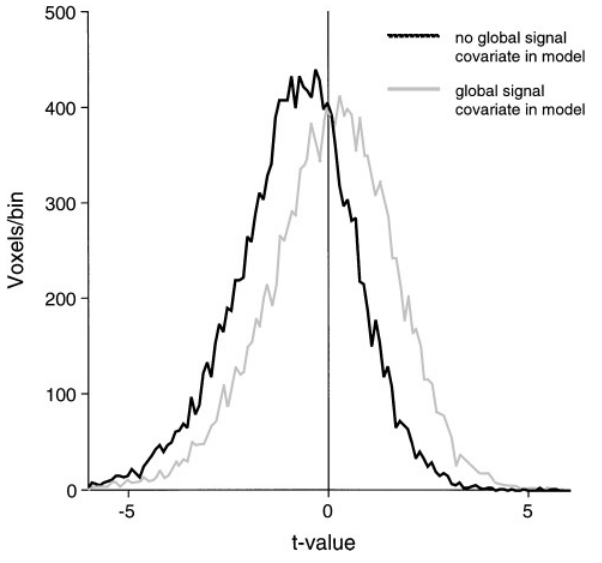

FIG. 7. Shown are two test statistic distributions (the statistical values corresponding to the assumed behavioral paradigm) from a representative human noise dataset which were generated with different regression models. One regression model contained a global signal covariate (gray) while the other did not (black). The regression models in this case assumed temporal independence. The distribution resulting from the regression model that included a global signal covariate is more centered about 0. This trend was also followed in the other human noise datasets, leading to a large reduction in the variance of distribution means across datasets.

served. In addition, simulations (data not shown) suggest that cardiac cycle aliasing would not manifest a 1/*f* component. Also, cardiac and respiratory effects in our data would not be expected to be great in magnitude as the pulse sequence employed is relatively insensitive to inflow effects (Jezzard and Song, 1996). Finally, as the 1/*f* component is present in data acquired from water phantoms, its presence apparently does not imply a physiological source (though physiological contributions cannot be ruled out).

The importance of lower frequencies (0.0083–0.025 Hz) to functional imaging is that they are in the bandwidth frequently used for behavioral paradigms. Due to the observed 1/*f* autocorrelation structure, statistical analyses of task effects in this bandwidth which assume independence will be biased toward rejecting the null hypothesis more often than expected. This was observed empirically. When independence was assumed, the *t* statistic distributions had broader tails than tabular distributions, as observed with FP1. The misestimation of the true α was also quite egregious when considering the nominal Bonferroni corrected significance of the maximum test statistic in each map as measured with FP2. The observed FP2 when independence was assumed was 59% compared to the theoretically expected 5%. Thus, the observed map-wise false-positive rate for these analyses was more than 10 times what an experimenter might expect if temporal independence was assumed. As statistical maps are generally thresholded and searched for excursion points to detect activations, this finding would seem quite relevant for fMRI. These results provide empirical evidence that performing a parametric statistical analysis that ignores temporal autocorrelation in BOLD fMRI data may result in rejecting the null hypothesis significantly more often than the rate specified by the nominal α, leading to unreliable results.

The autocorrelation seemed to be modeled well by a function in which power varies inversely with frequency. Even though there was a pattern of increasing power at lower frequencies in all datasets, the specific autocorrelation structure seemed to vary between datasets. This could be seen in the individual variability between power spectra in the 1/*f* components as well as in the flat frequency response components. We presented evidence that suggests the latter variability was caused by variations in noise emanating from stimulus presentation equipment throughout the course of data collection. Between-run variation in autocorrelation is sufficient to have caused the relatively large between-run variation in GLM false-positive rates (as evidenced by the greater variation in FP1 than expected if all of the datasets were from the same population). Since only one noise dataset was obtained per subject, it could not be determined if this variation was attributable to each subject or to each run.

In addition to the directly observable between-dataset variation in temporal autocorrelation, there was indirect evidence for within-dataset variation in temporal autocorrelation. This derives from the observed dissociation between FP1 and FP2 when the empirically based autocorrelation model was used as the sole model of autocorrelation in the modified GLM. A spatially nonstationary temporal autocorrelation structure could have resulted in an overestimation of relative power at the paradigm frequency for the majority of the voxels. Such an overestimation could have been sufficient to have caused the apparent overcorrection of FP1 whenever **K** included the empirically derived 1/*f* component. There has been evidence presented for tissue-specific autocorrelation with cortical regions displaying greater low frequency fluctuations than white matter (Weisskoff *et al.*, 1993; Biswal *et al.*, 1996a). This sort of spatially nonstationary temporal autocorrelation could have contributed to the observed dissociation in false-positive measures for the analysis using the empirically derived model of intrinsic autocorrelation as the sole component of **K**. Thus the employed model of intrinsic autocorrelation could have overspecified white matter autocorrelation (causing the low percentage of voxels greater than threshold, i.e., FP1) and underspecified cortical autocorrelation (causing the Bonferroni corrected threshold, i.e., FP2, to be superseded more often than expected by the FP1 val-

ues). A related observation is that the FP1 measure was less than theoretically expected in all of the analyses in which the **K** matrix included the empirically derived $1/f$ component. This further supports the idea that the empirically derived model of intrinsic autocorrelation utilized was not a valid description for all voxels.

The use of temporal smoothing kernels brought the false-positive rates closer to theoretically expected values compared to the assumed-independence case. However, the assumption that intrinsic autocorrelation would be negligible after exogenous smoothing was performed (Friston *et al.*, 1995c) was not given unqualified support as the specific exogenous smoothing kernel used had an effect on false-positive rates in fMRI noise data. When a Poisson kernel with parameter = 8 s was used (Friston *et al.*, 1994), group false-positive measures were not different from theoretically expected. However, when our empirically derived, narrower time-width smoothing kernel was used, FP1 was slightly, but significantly, higher than tabular values, while FP2 had a trend toward significance. Thus the validity of the analysis seemed to depend to some extent upon the bandwidth characteristics of the exogenous smoothing kernel. A possible explanation for this apparent filter-dependence is that the narrower time-width kernel passes more high frequencies than the wider time-width kernel, thus the analysis utilizing it relied on a larger part of the spectrum for error variance estimation. This analysis was in turn more sensitive to the correct specification of the intrinsic temporal autocorrelation structure of the data at these higher frequencies than the model that used the lower-passing Poisson kernel.

The observed beneficial effect of temporal smoothing on the agreement between the empirically observed and the theoretically expected false-positive rates can be explained by the stabilizing effect of smoothing on the temporal autocorrelation structure. This stabilizing effect renders assumptions about the stationality of the temporal autocorrelation structure across voxels and the negligible effect of intrinsic correlations more appropriate. We conclude that low-pass filtering (Worsley and Friston, 1995) can improve the agreement of observed with expected false-positive rates, as expressed by Friston *et al.* (1995c).

Biswal *et al.* (1996a) investigated cardiac and respiratory artifacts in noise fMRI time series, though they also observed lower frequency components. They showed that filtering targeted at these components could increase sensitivity to activation more than random filtering. The major source of nonwhite noise we have systematically observed is low frequency (primarily <0.05 Hz). This is too low in frequency to be respiratory (typically about 0.13 Hz), and we have seen that cardiac rates of about 1 Hz would not alias strongly to this range nor with a $1/f$ pattern (simulations, data not shown). Thus this low-frequency noise does not seem to be attributable to cardiac or respiratory effects. The problem caused in that report by cardiac and respiratory effects was to increase noise at frequencies other than typical paradigm frequencies, thus reducing sensitivity (by inflating the error variance; Biswal *et al.*, 1996a). The problem caused in the current report by the greater low-frequency noise was that it caused the observed false-positive rate to exceed that which was expected when the paradigm was in that range of frequencies, thus reducing specificity. Weisskoff *et al.* (1993) also reported the presence of low-frequency fluctuations in noise fMRI data, as well as in cardiac and respiratory effects. We did not attempt to filter respiratory effects because we did not generally observe, nor expect (Jezzard and Song, 1996) them in our data at a relative magnitude comparable to that reported by Biswal *et al.* (1996a).

To maximize sensitivity to activations, it has been suggested that the choice of temporal smoothing filter should mirror the IRF of the BOLD system (Friston *et al.*, 1995c). One exogenous smoothing filter used in this study, our empirically determined putative BOLD IRF, was obtained from actual activations observed in datasets in which subjects were engaged in a behavioral paradigm. The other putative BOLD IRF, a Poisson (parameter = 8 s), was obtained from a model-driven approach that examined intrinsic autocorrelations in the residuals of fMRI data (Friston *et al.*, 1994). This Poisson is narrower band than the BOLD IRF estimate obtained with our method. Another group has observed a putative BOLD IRF similar to the one obtained here by using different methods that also involve BOLD fMRI signal responses (Boynton *et al.*, 1996). The analyses presented here only dealt with null-hypothesis or noise data, and thus the issue of optimal temporal smoothing could not be addressed (however, see Aguirre *et al.*, 1997). Additionally, as the BOLD system does not seem to be perfectly linear (Vasquez and Noll, 1996; Boynton *et al.*, 1996), it does not actually posses a true IRF. Thus, the term "IRF" is being used in this report in an informal way that refers to the near linearity of the BOLD system (Vasquez and Noll, 1996; Boynton *et al.*, 1996).

By examining global signals, it was observed that spatially unsmoothed fMRI datasets, both human noise and phantom, are not composed of voxel time series which are uncorrelated in space, but rather have spatial coherency apparently extending to all temporal frequencies (we collected data from 0.003125–0.025 Hz). However, the low temporal frequencies are much more spatially coherent than the higher frequencies. The inclusion of a global signal covariate in the GLM stabilized (i.e., reduced the variance of) the dataset false-positive rates. The reason for this is that by including a global signal covariate, the voxel time

series within a dataset were made more spatially independent, thus reducing the shifts of the means of the test-statistic distributions about 0. This simultaneously reduced the variation across datasets in the extent to which the tails of the distributions extended beyond threshold (i.e., the FP1 measure). It would thus seem that beyond the possibility of significantly reducing spatially coherent noise in fMRI time series, the inclusion of a global signal covariate might also improve the validity of voxel-wise statistical analysis on a dataset-wise basis in spatially unsmoothed data. However, the effects of including a global signal covariate become more complex in a behavioral dataset where activation is a possibility (see Aguirre *et al.*, 1997).

The cause(s) of the spatial coherency of the fMRI global signal is not understood. The global signal seems to contain components which are not attributable to at least the first-order effects of motion. Regionally coherent low-frequency signal components have been observed (Biswal *et al.*, 1995) and may bear some mechanistic relation to the global signal observed here. A physiological basis of these regional signals has been suggested by their absence during hypercapnia (Biswal *et al.*, 1996b). The spatial pattern of the global signal in space did not immediately suggest a mechanism to the authors. It was present at a relatively low level in a large part of the datasets examined, with patches of higher correlation being present throughout the brain. In theory, smoothness due to a stationary, continuously differentiable spatial autocovariance function could have yielded the observed spatial coherency. However, examinations of smoothed, computer-generated noise suggested that this is not the case. Perhaps more sophisticated multivariate techniques such as principal components analysis could also be used to help better understand the source(s) of the spatially coherent signals in null-hypothesis data. In the same vein, the presence of such strong spatial coherencies (especially at paradigm frequencies) in null-hypothesis data should be a caveat to the use of multivariate methods to detect "activations" in a paradigm-free manner.

CONCLUSIONS

It was found that:

(1) Temporal autocorrelation in BOLD fMRI "noise" data (i.e., data collected under the null hypothesis) in both humans and phantoms is characterized reasonably by a frequency domain model that includes a $1/f$ component. This $1/f$ component could not be completely explained by the first-order effects of motion, nor could it be explained by other equipment present during scanning, nor could it be explained by convolution of neural activity by hemodynamics (Friston *et al.*, 1994).

(2) When independence was assumed in statistical analyses of human noise datasets that assumed a square wave input at 0.0125 Hz (i.e., 40 s off, 40 s on), observed false-positive rates significantly exceeded expected false-positive rates.

(3) When a modified GLM that takes temporal autocorrelation into account was used (Worsley and Friston, 1995), an empirically derived model of intrinsic temporal autocorrelation (using averages of power spectra across voxels and datasets) for BOLD fMRI data proved to be invalid. This suggests a spatially nonstationary temporal autocorrelation structure.

(4) Temporal smoothing of the noise data with low-pass filters in conjunction with the use of the modified GLM (Worsley and Friston, 1995) resulted in false-positive rates in null-hypothesis data that were close to theoretically expected results. However, there was some filter dependence on the agreement with the theoretically expected results.

(5) fMRI datasets evidenced spatial coherency (i.e., dependence between voxel time series). The spatial coherency was greater at lower temporal frequencies. This spatial coherency could not be explained by a stationary, continuously differentiable autocorrelation function [akin to that used in the context of SPM statistics (Friston *et al.*, 1991; Worsley *et al.*, 1992)].

(6) Inclusion of global signal covariates in the regression models tended to center the test-statistic distributions of individual datasets around 0 and stabilized (i.e., reduced the variance of) voxel-wise false-positive rates across datasets.

The source(s) of the spatial coherency that was observed in fMRI datasets is, as yet, undetermined. Paradigm-free, multivariate analysis techniques such as fuzzy-clustering, principal components analysis, and others assume that the expected spatial coherency is 0 under the null hypothesis. How the observed spatial coherency should impact the interpretation of the results of these techniques would seem to be an important issue.

Though other groups have reported observations of power spectra that appear to have increasing power at lower frequencies (Weisskoff *et al.*, 1993; Friston *et al.*, 1994; Boynton *et al.*, 1996; Jezzard and Song, 1996), a quantitative comparison between labs has not yet been performed. Until such a comparison has been made, it would not be correct to assume that the expected temporal autocorrelation is the same in every scanner/site, let alone for every fMRI scanning method. However, based on the observations reported here, it would appear to be highly recommended, and productive, for statistical analyses to be validated on a sample of noise data. It should also be noted that most nonparametric as well as parametric statistics assume independence. Thus groups that prefer to use nonparametric statistics would also benefit from empirical characterization of

the temporal autocorrelation found in their data and its effects on voxel-wise statistics.

The use of noise datasets should be contrasted with the use of computer-generated noise simulations which may not possess the characteristics of actual data. The use of noise datasets should also be contrasted with attempts at specificity validations that involve performing analyses on "shuffled" data from experimental datasets. Such shuffling causes a loss of temporal information, and thus effects of nonindependence (like the ones observed in the current report) become undetectable. In contrast, by using datasets that lack experimental treatments, one should obtain valid estimates of specificity.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (NS01762 and AG13483), the McDonnell–Pew Program in Cognitive Neuroscience, and the Charles A. Dana Foundation.

REFERENCES

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

Bandettini, P. A. 1993. Studies of brain activation: Temporal characteristics. In *Functional MRI of the Brain,* pp. 143–151. *Soc. Magn. Reson. Med.*, Berkeley, CA.

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

Biswal, B., DeYoe, E. A., and Hyde, J. S. 1996a. Reduction of physiological fluctuations in fMRI using digital filters. *Magn. Reson. Med.* **35:**107–113.

Biswal, B., Hudetz, A. G., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. 1996b. Hypercapnia reversibly diminishes low-frequency fluctuations in human motor cortex using EPI. *Proc. Int. Soc. Magn. Reson. Med.* **31:**1760. [Abstract]

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:**537–541.

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:**4207–4221.

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

Friston, K. J., Frith, C. D., Liddle, P. F. Dolan, R. J., Lammertsma, A. A., and Frackowiak, R. S. 1990. The relationship between global and local changes in PET scans. *J. Cereb. Blood Flow Metab.* **10:** 458–466.

Friston, K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. S. J. 1991. Comparing functional (PET) images: The assessment of significant change. *J. Cereb. Blood Flow Metab.* **11:**690–699.

Friston, K. J., Frith, C. D., Turner, R., and Frackowiak, R. S. J. 1995b. Characterizing evoked hemodynamics with fMRI. *NeuroImage* **2:**157–165.

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

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-B., Frith, C. D., and Frackowiak, R. S. J. 1995d. Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Map.* **2:**189–210.

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

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

Haxby, J., Grady, C., Horwitz, B., Ungerleider, L., Mishkin, M., Carson, R., Herscovitch, P., Shapiro, M., and Rapoport, S. 1991. Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proc. Natl. Acad. Sci. USA* **88:**1621–1625.

Holmes, A. P. 1994. *Statistical Issues in Functional Brain Mapping.* Doctor of Philosophy thesis, University of Glasgow.

Jezzard, P. 1996. Effects of static field drift on echo planar functional MRI. *NeuroImage* **3:**S30. [Abstract]

Jezzard, P., and Song, A. W. 1996. Technical foundations and pitfalls of clinical fMRI. *NeuroImage* **4:**S63–S75.

Jiang, A., Kennedy, D. N., Baker, J. R., Weisskoff, R. M., Tootell, B. H., Woods, R. P., Benson, R. R., Kwong, K. K., Brady, T. J., Rosen, B. R., and Belliveau, J. W. 1995. Motion detection and correction in functional MR imaging. *Hum. Brain Map.* **3:**224–235.

Kwong, K. K., Beliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R., *et al.* 1992. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Nat. Acad. Sci. USA* **89:**5675–5679.

Ogawa, S., Menon, R. S., Tank, D. W., Kim, S. G., Merkle, H., Ellermann, J. M., and Ugurbil, K. 1993. Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging: A comparison of signal characteristics with a biophysical model. *Biophys. J.* **64:**803–812.

Oppenheim, A. V., Willsky, A. S., and Young, I. T. 1983. *Signals and Systems.* Prentice–Hall, Englewood Cliffs, NJ.

Shin, R. K., D'Esposito, M., Detre, J. A., Alsop, D. C., Atlas, S. W., and Grossman, M. 1995. Functional magnetic resonance imaging activation during a visuospatial/mental rotation task. *Neurology* **45:** A267. [Abstract presented at the American Academy of Neurology Meeting, Seattle, WA]

Vasquez, A. L., and Noll, D. C. 1996. Non-linear temporal aspects of the BOLD response in fMRI. *Proc. Soc. Magn. Reson. Med.* **3:**1765. [Abstract]

Weisskoff, R. M., Baker, J., Belliveau, J., Davis, T. L., Kwong, K. K., Cohen, M. S., and Rosen, B. R. 1993. Power spectrum analysis of functionally-weighted MR data: What's in the noise? *Proc. Soc. Magn. Reson. Med.* **1:**7. [Abstract]

Woods, R. P., Cherry, S. R., and Mazziotta, J. C. 1992. Rapid automated algorithm for aligning and reslicing PET images. *J. Comput. Assisted Tomogr.* **16:**620–633.

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

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

Xiong, J., Gao, J., Lancaster, J. L., and Fox, P. T. 1995. Clustered pixels analysis for functional MRI activation studies of the human brain. *Hum. Brain Map.* **3:**287–301.