Nonlinear Temporal Dynamics of the Cerebral Blood Flow Response

Karla L. Miller,[4] Wen-Ming Luh,[1] Thomas T. Liu,[1] Antigona Martinez,[1] Takayuki Obata,[1] Eric C. Wong,[1,2] Lawrence R. Frank,[1,3] and Richard B. Buxton[1]*

[1]*Department of Radiology, University of California at San Diego, San Diego, California*
[2]*Department of Psychiatry, University of California at San Diego, San Diego, California*
[3]*VA San Diego Healthcare System, San Diego, California*
[4]*Department of Electrical Engineering, Stanford University, Stanford, California*

Abstract: The linearity of the cerebral perfusion response relative to stimulus duration is an important consideration in the characterization of the relationship between regional cerebral blood flow (CBF), cerebral metabolism, and the blood oxygenation level dependent (BOLD) signal. It is also a critical component in the design and analysis of functional neuroimaging studies. To study the linearity of the CBF response to different duration stimuli, the perfusion response in primary motor and visual cortices was measured during stimulation using an arterial spin labeling technique with magnetic resonance imaging (MRI) that allows simultaneous measurement of CBF and BOLD changes. In each study, the perfusion response was measured for stimuli lasting 2, 6, and 18 sec. The CBF response was found in general to be nonlinearly related to stimulus duration, although the strength of nonlinearity varied between the motor and visual cortices. In contrast, the BOLD response was found to be strongly nonlinear in both regions studied, in agreement with previous findings. The observed nonlinearities are consistent with a model with a nonlinear step from stimulus to neural activity, a linear step from neural activity to CBF change, and a nonlinear step from CBF change to BOLD signal change. *Hum. Brain Mapping 13:1–12, 2001.* © 2001 Wiley-Liss, Inc.

Key words: functional MRI; cerebral blood flow; arterial spin labeling; nonlinear dynamics; hemodynamics

INTRODUCTION

Functional magnetic resonance imaging (fMRI) detects activation in the brain by measuring changes in the MR signal during neural stimulation. The most common technique, blood oxygenation level dependent (BOLD) fMRI, is sensitive to changes in the concentration of deoxyhemoglobin in small local blood vessels [Kwong et al., 1992; Ogawa et al., 1992; Bandettini et al., 1992]. A decrease in deoxyhemoglobin concentration follows activation in response to a disproportionately large increase in cerebral blood flow (CBF) accompanying a smaller increase in the cerebral metabolic rate of oxygen ($CMRO_2$) [Fox and Raichle, 1986; Fox et al., 1988]. However, the details of these dynamics are complicated and poorly understood, involving cerebral blood volume (CBV) in addition to CBF and $CMRO_2$. In this work, we consider these physiological responses to arise in the pathway dia-

Contract grant sponsor: NINDS-36722, VA SA 321 (Dr. Frank).

*Correspondence to: Richard B. Buxton, Department of Radiology, 8756, UCSD Medical Center, 200 West Arbor Drive, San Diego, CA 92103. E-mail: rbuxton@ucsd.edu

Received for publication 28 April 2000; accepted 13 December 2000

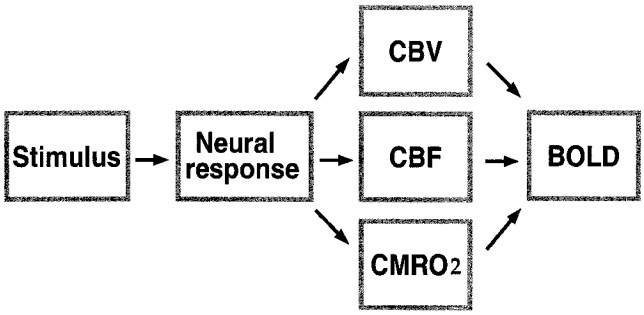

Figure 1.
Diagram of the transformation from stimulus presentation to BOLD response. At each step in the model, nonlinearities may be introduced.

grammed in Figure 1. Recent advances in perfusion imaging with arterial spin labeling (ASL) enable increasingly accurate measurement of the timecourse of local CBF during neural activity [Kim, 1995; Ye et al., 1997; Wong et al., 1998]. Simultaneous measurement of the CBF and BOLD responses to a stimulus allows us to begin to explore the connections sketched in Figure 1.

An important aspect of these dynamics concerns which steps in the transformation from stimulus to measured response are linear and time invariant. Under such a relationship, the output physiological response would (1) exhibit the linear property of superposition (i.e., a weighted sum of input waveforms would produce a weighted sum of the individual output responses) and (2) be independent of input timing (i.e., a time shift in the input would cause the same time shift in the output response). Steps that are linear and time invariant would be simple to describe and would exhibit a number of convenient properties. In addition to modeling considerations, these characteristics have important implications for a number of experimental paradigms (such as event-related fMRI) [Dale and Buckner, 1997; Liu and Gao, 1999; Duyn et al., 2000; Yang et al., 2000] and data processing schemes (such as correlation analysis) [Bandettini et al., 1993], which assume that fMRI signal dynamics bear a linear, time-invariant relationship to the stimulus. Specifically, the BOLD response is often modeled as a convolution of the stimulus pattern with a smooth hemodynamic impulse response function (which may vary spatially [Lange and Zeger, 1997]), such as a Gaussian or gamma-variate function. The usefulness of this type of model makes a detailed characterization of the relationship between stimulus and response worthwhile. Even if the hemodynamic processes of interest do not linearly follow stimulus presentation, a linear approximation may be quite accurate in certain circumstances [Friston et al., 1998; Glover, 1999].

Previous studies have demonstrated a nonlinear relationship between stimulus and BOLD response, in terms of both stimulus duration and magnitude. The BOLD response to short-duration stimuli cannot in general be used to linearly predict the response to long-duration stimuli in the visual [Boynton et al., 1996; Vazquez and Noll, 1998; Hoge et al., 1999], auditory [Robson et al., 1998; Glover, 1999], or motor [Glover, 1999] cortices. However, some of these studies also report that the BOLD response begins to behave linearly when the short-duration response used for the prediction is longer than some threshold duration, estimated at 4–6 sec in the visual [Boynton et al., 1996; Vazquez and Noll, 1998] and auditory cortices [Robson et al., 1998]. Other work has examined the linearity of the BOLD response to stimulus magnitude, finding BOLD to be nonlinearly related to visual stimulus contrast [Boynton et al., 1996; Vazquez and Noll, 1998] and word presentation rates [Rees et al., 1997; Buchel et al., 1998; Friston et al., 1998]. The nonlinearities observed have generally taken the form of overprediction of response magnitude and underprediction of response duration. These nonlinearities in the BOLD response suggest that care should be taken in applying techniques that assume a linear relationship between stimulus and BOLD response.

The linearity of the perfusion (CBF) response is, to date, not as well studied as the BOLD response. Previous work has produced conflicting results concerning the linearity of the CBF response in different brain regions, with a PET study finding a linear relationship between word presentation rate and perfusion [Rees et al., 1997] and ASL perfusion studies finding a nonlinear relationship between visual stimulus contrast and perfusion [Hoge et al., 1999; Yang et al., 2000]. Because changes in CBF are only one component of the BOLD effect, we do not necessarily expect to find the same nonlinearities in the CBF response as in the BOLD response. Simultaneous measurement of these two responses allows us to separate BOLD nonlinearities that arise in the steps between stimulus and CBF from those that are introduced in the transition from flow response to BOLD response.

In this article, we consider the transformations from the stimulus to the perfusion and BOLD responses by studying the effect of varying stimulus duration. Our goal is to characterize these dynamics using a single experimental paradigm in two areas of the brain. We consider whether the perfusion response is linearly related to the stimulus, verify previous findings of the nonlinear relationship between stimulus and BOLD

effect, and discuss the implications of our observations in terms of the model shown in Figure 1. This study is motivated by an interest in clarifying the differences between BOLD and perfusion fMRI in order to better understand the role that perfusion techniques may play in functional neuroimaging. Preliminary reports on this work were presented previously [Miller et al., 1999, 2000].

EXPERIMENTAL METHODS

Stimulus presentation

All experiments presented here follow the same stimulus presentation design. The initial 20 and final 10 sec of each presentation run were without stimulus. The remainder of the run consisted of eight identical cycles of stimulus presentation for some fixed duration (2, 6, or 18 sec) followed by 19 sec without stimulus. For each subject, two runs were collected for each stimulus duration, for a total of 16 cycles of presentation at each duration. An additional run of the 18-sec stimulus was collected on each subject for the sole purpose of selection of activated voxels.

The above experimental design was used in separate experiments with two stimulus types, chosen to elicit a strong response from either the primary visual or primary motor cortex. The visual stimulus was a black-and-white radial checkerboard pattern flickering at 8 Hz and projected onto a screen inside the scanner bore. The visual stimulus during rest periods was a small white fixation cross at the center of a black visual field. The motor stimulus consisted of sequential bilateral finger tapping cued by the occurrence of the visual stimulus from the visual experiment. Subjects were instructed to tap at a constant, comfortable rate. Visual stimuli were presented using PsyScope [Cohen et al., 1993]. The visual experiment was performed on three subjects and the motor experiment was performed on four subjects. Informed consent was obtained from all subjects in accordance with local institutional review board guidelines.

Data acquisition

ASL data was collected in each experiment using an echo-planar QUIPSS II [Wong et al., 1998] pulse sequence (TR = 2,000 ms, TE = 30 ms, TI1 = 700 ms, TI2 = 1,400 ms) with 8-mm thick slices. This technique gathers images in alternating conditions of tag (magnetic tagging of arterial blood) and control (no tagging). The perfusion signal is the difference between tag and control images, while the BOLD signal is encoded in the average of tag and control conditions [Wong et al., 1997]. Because BOLD is an additive signal (control + tag) and flow is a difference signal (control − tag), a simultaneous but independent measurement of the two signal time series can be constructed from the same data set (assuming independent Gaussian noise).

Data for the motor experiment was collected on a GE Signa 1.5 T scanner fitted with a local gradient head coil [Wong et al., 1992]. A PICORE QUIPSS II sequence captured a field of view of 24 cm with 3.75×3.75 mm^2 in-plane resolution. Three contiguous 8-mm thick axial slices through the primary motor cortex were acquired, although only the most activated slice for each subject was included in data analysis. Data for the visual experiments was collected on a Siemens Vision 1.5 T scanner with a receive-only flexible surface coil centered on the occipital cortex. This setup enables tagging of the arterial blood to be performed with the transmit body coil, improving the quality of the tag over a gradient head coil. An EPISTAR QUIPSS II sequence captured a field of view of 48 cm with 3.75×3.75 mm^2 in-plane resolution, which was later windowed down to a 24-cm field of view (images being acquired at the larger field of view to improve the signal-to-noise ratio). The visual experiment acquired an 8-mm thick oblique axial slice along the calcarine fissure. The use of different scanning hardware for the two experiments is not expected to affect our results, although further experiments to test this would be worthwhile.

Data processing

Regions of interest (ROIs) were selected from the additional 18-sec stimulus run, which was not included in the final data. The use of a separate run enables the identification of activated voxels independent of the data used in later analyses, eliminating bias in the comparison of the responses to stimuli of different duration. Because our primary goal was to characterize the CBF response and because the largest BOLD signals tend to originate in draining veins where there are only minor perfusion changes [Buxton et al., 1997], voxels were selected based on the flow signal in the additional run. This run was collapsed into a single cycle of stimulus response by averaging the measured raw data timecourses for each stimulus cycle. The flow signal was then calculated from this average cycle by subtracting each image from the average of the image before and image after. This single cycle of flow was correlated with a trapezoidal stimulus reference function at each voxel [Bandettini

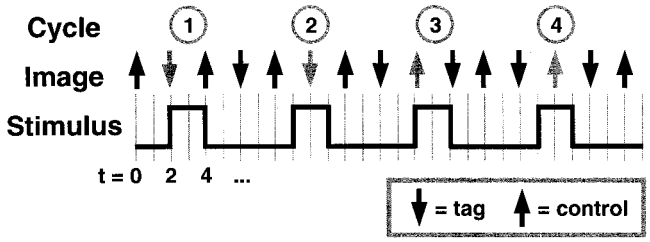

Figure 2.

Timing of imaging relative to stimulus presentation. Images are acquired every 2 sec ($TR = 2$ sec), alternating between control and tag conditions. The presented stimulus is repeated with a period of $2n + 1$ seconds ($n = 3$ above) so that over four consecutive stimulus cycles, a control and a tag image are acquired at each second of the stimulus cycle. The gray arrows illustrate tag and control images that are gathered with 1-sec resolution over a 2-sec window of the stimulus cycle.

et al., 1993] and activated voxels were identified by thresholding the correlation coefficients (where the threshold ranged from 0.55 to 0.85, chosen for each subject as the correlation coefficient at which the identified voxels were confined to the cortical region of interest).

Images were spatially registered using AFNI software [Cox, 1996] and the CBF and BOLD responses to single cycles of stimulus were calculated as follows. Because the length of each stimulus cycle (an odd number of seconds) was asynchronous with image acquisition (every 2 sec), the signal over four consecutive stimulus cycles contains a control and a tag image at each second of the stimulus cycle (see Fig. 2). A single short run of four consecutive stimulus cycles was generated by averaging the two identical runs of eight cycles together and then averaging the first and second halves of the resulting average. Single-cycle BOLD and flow responses were then generated with 1-sec time resolution, BOLD as the average signal (control + tag) and flow as the difference signal (control − tag) at each time point.

This calculation of the flow signal interleaves the measurements acquired over different stimulus cycles, allowing low frequency effects in the raw signal (e.g., amplitude or baseline drifts) to alias to high frequencies in the flow signal. The subtraction of a fourth-order polynomial from each pixel timecourse immediately after registration was able to remove significant baseline drifts without interfering with the 8-cycle stimulus pattern. In addition, variation of the BOLD response in different cycles of stimulus presentation will also produce low frequency effects that could alias into the calculated flow signal. Because the flow signal is much smaller than the raw signal, these aliasing effects can be significant. Complete removal of this artifact would require significant smoothing, degrading the temporal resolution of the response curves to 4 sec or more. Rather than increase the temporal resolution by such a large amount, we applied a smoothing filter with a full width at half maximum (FWHM) of 2 sec, which did not completely remove the aliased artifact but did make it less prominent. A single flow and BOLD curve for each stimulus type and duration were calculated for each subject by averaging the smoothed single-cycle response in activated voxels.

Linearity analysis

In order for a physiological response to be a linear and time-invariant transformation of the stimulus, the response to a particular stimulus must be described by the convolution of the stimulus presentation pattern with the response to a very brief stimulus (for sufficiently short stimuli, the latter response is called the *impulse response*). In other words, one requirement of linearity is that the response to a long stimulus should be equal to an appropriately shifted and summed series of short responses. We tested the linearity of our data by summing shifted replicas of the response to short stimuli to match the duration of longer stimuli. For example, the 2-sec response was replicated three times, shifted by 0, 2, and 4 sec, and then summed. This summation was then compared to the observed response to the 6-sec stimulus. Our experimental paradigm enabled three such comparisons: in addition to the above comparison of the 2- and 6-sec responses, we compared the 2- and 18-sec responses and the 6- and 18-sec responses.

In a linear system, all order of moments of the linearly predicted response and measured response should be equal. Accordingly, we can establish statistically significant departure from linearity by showing, for any order of moment, a significant difference between the predicted response moment and the measured response moment. For example, if the second-order moment of the predicted and measured responses is significantly different, the response is nonlinear even if lower-order moments do not differ. The fractional moment difference of a predicted response p(t) and a measured response m(t) was calculated for each of the linearity comparisons as:

$$M_n = \frac{\int t^n p(t)dt - \int t^n m(t)dt}{\int t^n p(t)dt} \qquad (1)$$

If M_n is significantly different from zero, the n^{th} moment of the predicted response does not equal the n^{th}

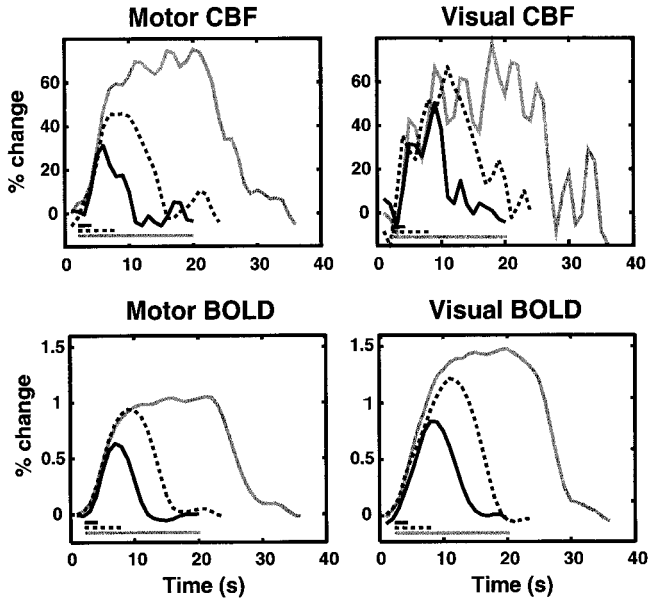

Figure 3.
Average flow and BOLD signal measurements. Each experiment consisted of 16 repetitions of three durations of motor or visual stimulus (2, 6, and 18 sec) followed by a resting period with no stimulus (19 sec). The measurements shown are the intersubject average response to a single cycle of stimulus. Responses are expressed as percent signal change from baseline.

moment of the measured response, and the system in question is nonlinear.

RESULTS

The intersubject average curves for the different stimuli are shown in Figure 3. Note that the high-frequency ripple in the flow signal is caused by the raw signal aliasing artifact mentioned earlier, and not noise. The comparisons of the predicted versus measured curves are shown for the flow response in Figure 4 and for the BOLD response in Figure 5.

Although a linear model of flow tended to overpredict the measured response (Fig. 4), the extent of overprediction was not consistent across the visual and motor areas. In the visual cortex, a strong overprediction of the response magnitude change was observed for all three comparisons performed. This overprediction ranged from about 1.5× the measured response (for the prediction of the 18-sec response by the 6-sec response) to about 2.4× (for the prediction of the 18-sec response by the 2-sec response). In the motor cortex, however, only a small overprediction of about 1.3× was observed when the 2-sec response was used to predict the 6-sec response, and both the 2- and 6-sec responses predict the amplitude of the 18-sec response reasonably well.

Our linearity analysis of the BOLD response (Fig. 5) found that the linear model consistently overpredicted

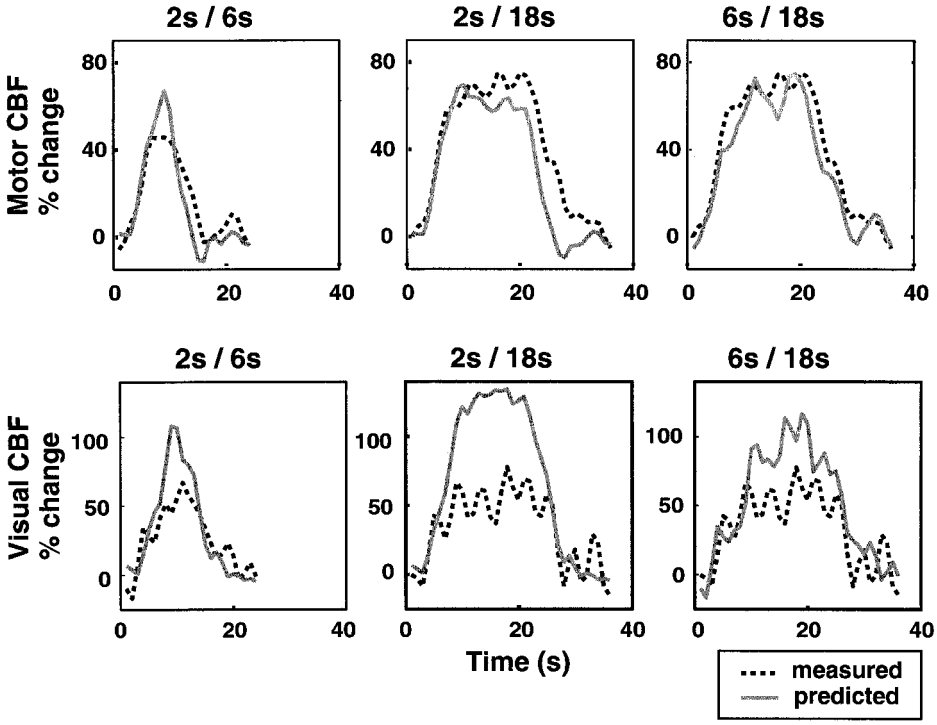

Figure 4.
Linear, time-invariance analysis of motor and visual flow data shown in Figure 3. Plot titles indicate the duration of the short- and long-duration responses used in each analysis. Linearity was tested by appropriately replicating, shifting, and summing the measured response to a short-duration stimulus. If flow is a linear transformation of the stimulus, this superposition (solid gray lines) should predict the measured response to the long-duration stimulus (dotted black lines). A mismatch of the predicted and measured long-duration responses indicates a nonlinear response. The flow response in the motor cortex appears to be nearly linear (with a slight overprediction of the 6-sec response), whereas the flow response in the visual cortex behaves nonlinearly.

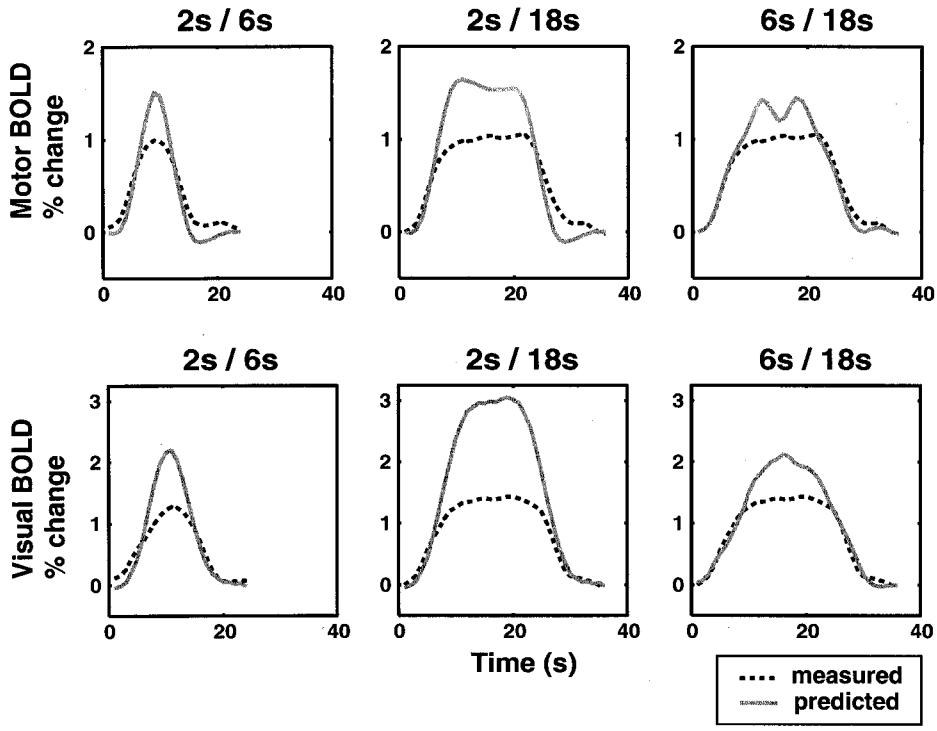

Figure 5.
Linear, time-invariance analysis of motor and visual BOLD data shown in Figure 3. Plot titles indicate the duration of the short- and long-duration responses used in each analysis. Linearity was tested by appropriately replicating, shifting, and summing the measured response to a short-duration stimulus as in Figure 4. The BOLD response in both the visual and motor cortices is nonlinearly related to stimulus duration.

the peak response magnitude in both the visual and motor cortices, with the motor cortex showing somewhat smaller overpredictions (1.3–1.6× the measured response) than the visual cortex (1.4–2.0×). In both cortices, the smallest overprediction occurred when the 6-sec response was used to predict the 18-sec response. Previous studies have reported that BOLD begins to conform to the linear model in this range [Boynton et al., 1996; Robson et al., 1998; Vazquez and Noll, 1998]. However, the data presented in each of these studies shows evidence for subtle nonlinear effects. Although our 6-sec response does not predict the 18-sec response with a linear model, it is a better linear predictor than the 2-sec response in the sense that the amount of overprediction is less severe.

The fractional moment difference M_n was calculated for n = 0–5 for each of the linearity comparisons shown in Figures 4 and 5 using individual subject data. The mean and standard deviation of M_n were then calculated for each linearity comparison. Figure 6 shows the mean (±1 SD) of M_n for the first three moments (moments 3–5 are not shown, as these results are similar to M_1 and M_2). None of the calculated fractional moment differences of the motor flow response deviate significantly from linearity. In contrast, the visual flow response shows a significant departure from nonlinearity, which is manifested as an overprediction of moment magnitude. The BOLD response is significantly nonlinear for all comparisons except the 2-sec prediction of the 6-sec motor response. Like the visual flow response, the predicted moments for the BOLD response are significantly larger than the measured moment.

DISCUSSION

In contrast to the BOLD response, which has consistent nonlinearities in both the primary visual and motor cortices, the nonlinearity of the flow response differs between the two regions, with a nearly linear motor flow response but a strong nonlinear visual flow response. Both the flow and BOLD response nonlinearities take the form of an overprediction of the response magnitude of long duration responses by short duration responses. To investigate possible sources of nonlinearities, we consider the pathway from stimulus to BOLD introduced earlier (Fig. 1). We would like to consider at which of these steps the observed nonlinearities are introduced. Although the measured BOLD responses in the two studied regions exhibited similar nonlinearities, our response pathway attributes these nonlinearities to different sources. The flow response to the motor stimulus was nearly linear, indicating that most of the nonlinearities observed in the motor BOLD signal change were introduced in the transformation from the flow response to BOLD response. In contrast, the flow response to the visual stimulus had strong nonlinearities similar to the

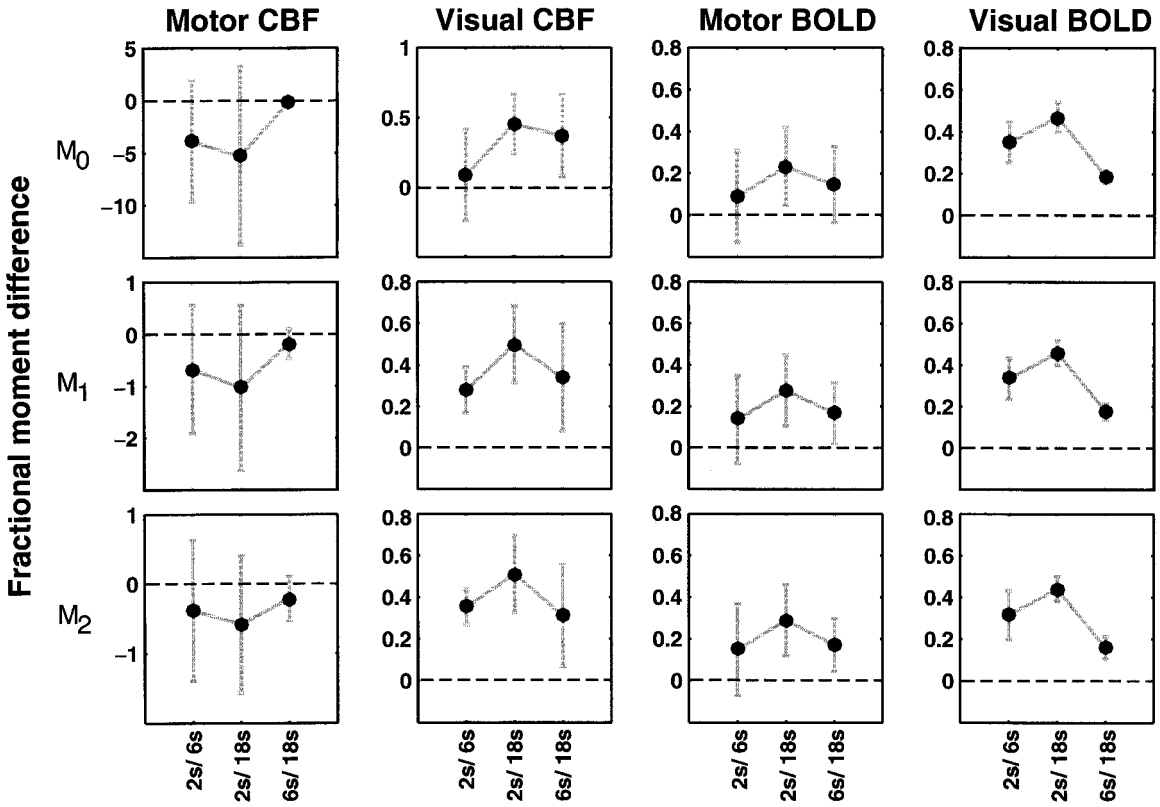

Figure 6.

Fractional moment differences (mean ± 1 SD) calculated according to Eq. 1 for the zeroth, first, and second moments (top to bottom) from individual subject data. Intersubject averages of the linearity comparisons are shown in Figures 4 and 5. A response is nonlinear if M_n is significantly different from zero for any *n*. The motor CBF response is not significantly nonlinear, but the visual CBF and BOLD responses are in general nonlinear.

BOLD response, indicating that most of the BOLD nonlinearities were already present in the perfusion response. Hence, nonlinearities appear to arise both in the steps from stimulus presentation to flow response and from flow response to BOLD effect. We consider whether specific nonlinearities that are known to be introduced in these steps are sufficient to describe the nonlinearities found in our data.

The perfusion response

Our goal in fMRI experiments is to observe the neural response to a stimulus, which is known in general to be a nonlinear transformation of the stimulus. A linear relationship between neural activity and flow response could therefore exist such that all nonlinearities in the flow response are already present in the neural response. Such a scenario would indicate that flow measurements provide a fairly simple way to accurately observe the (nonlinear) neural response to a stimulus. We wish to consider whether such a relationship is consistent with our results, given what is known about nonlinearities in neural activity.

For the first step of the response pathway from stimulus to neural response, we model known nonlinearities involving neural adaptation. Neural adaptation in this case refers to the general tendency of the neural firing rate to decay from a high initial value to a lower steady-state value during a sustained stimulus [Rieke et al., 1997]. This classic pattern has been observed in numerous recordings of cell firing rates [e.g., Albrecht et al., 1984; Maddess et al., 1988; Bonds, 1991]. Neural adaptation has previously been proposed as a source of BOLD nonlinearities in the visual cortex [Boynton et al., 1996] and in the auditory cortex [Robson et al., 1998]. Our parameterization of the neural response to a block stimulus that begins at $t = 0$ and ends at $t = T$ is:

$$n(t) = \begin{cases} 1 + ae^{-t/\tau_n} & 0 \le t < T \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

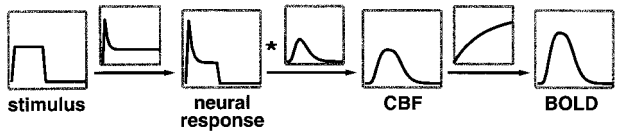

Figure 7.

Model for the steps from block stimulus presentation to BOLD response (based on the pathway introduced in Fig. 1). The neural response is modeled as a nonlinear function of the presented stimulus in which an initial strong response decays to a lower steady-state value (Eq. 2). This neural response is convolved with a flow impulse response model (the gamma-variate function of Eq. 3) to give the CBF response. This model for the CBF response is a nonlinear function of the stimulus, but a linear transformation of the neural response. Finally, the step from CBF to BOLD is described by a scaling curve that saturates with increasing flow.

with parameters τ_n, a decay time constant, and a, the amount that the initial response overshoots the steady-state response. This model has a disproportionately large response to short stimuli because of the transient overshoot of the neural firing rate following stimulus onset. Note that although this model is nonlinear in general, it is a linear transformation of the stimulus when a = 0, representing a neural response that exactly follows the input stimulus pattern. Although we have no measurements of the neural response, we can fit our flow data simultaneously to this neural response model and a model for the transition from neural response to flow.

For the step from neural response to CBF, we consider a simple linear, time-invariant transformation in which the neural response is convolved with a gamma-variate function, a common model for the hemodynamic impulse response [e.g., Boynton et al., 1996]:

$$h(t) = \frac{c}{\tau_h m!}\left(\frac{t - t_d}{\tau_h}\right)^m e^{-(t-t_d)/\tau_h} \qquad (3)$$

with parameters τ_h, t_d, c and m. This flow impulse response model is simply a smoothing kernel that blurs the neural response. The blurring effect is controlled by the parameters τ_h, the time constant for the kernel's falling edge, and m, which affects the shape of the function (with larger values of m increasing the kernel's symmetry, making it look more Gaussian). The t_d and c parameters delay and scale the response but do not affect its shape.

Convolving the nonlinear neural response model (Eq. 2) with the hemodynamic impulse response (Eq. 3) yields a single model relating the stimulus to the flow, as shown in Figure 7. This two-step model for the flow response was simultaneously fit to the average 2-, 6-, and 18-sec motor and visual flow responses shown in Figure 3, yielding a best fit model for each brain area. The model fits and their parameters are shown in Figure 8, and the model responses are shown in Figure 9. The model response curves for each region are generated by varying the input stimulus duration (T) for a fixed set of model parameters. These fits demonstrate that our flow data is consistent with a linear, time-invariant transformation of an adaptive neural response model.

The flexibility of this model (which has six parameters) enables a number of parameter sets to give a good fit to our data. In this fitting, our goal was to test whether variations in the parameter a with similar impulse response functions could account for the observed data. For this reason, a was the primary parameter that was varied, and the remaining parameters fine-tuned the shape of the response models. We found that a large difference in a (3.0 in visual and 0.25 in motor) combined with similar impulse response functions (FWHM of 6.2 sec in visual and 5.2 sec in motor) provides a reasonable account for the data. Recall that the a parameter determines the degree of adaptation exhibited by the neural response. A large initial overshoot (strong adaptation) will cause the

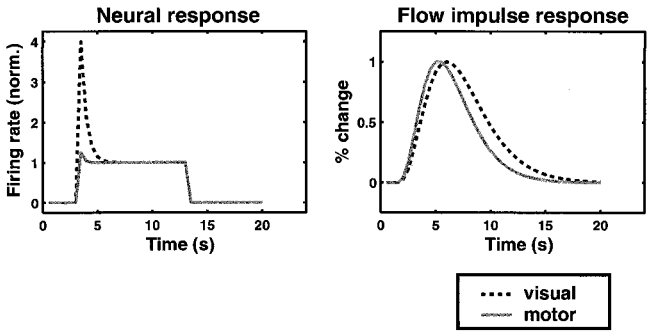

Figure 8.

Model functions fit to the motor and visual flow data shown in Figure 3. These functions describe the steps from stimulus to neural response (left subplot) and neural response to flow response (right subplot). The convolution of these two forms (described by Eqs. 2 and 3) were simultaneously fit to the measured flow response to 2, 6, and 18-sec stimuli in the motor area (τ_n = 0.25, t_d = 0.65, a = 0.25, τ_h = 1.25, c = 65, m = 3) and the visual area (τ_n = 0.5, t_d = 1.5, a = 3.0, τ_h = 1.5, c = 54, m = 3). Of the parameters that effect the functional form of the model (a, τ_n, m, and τ_h), only a differs significantly across the two fits. The (nearly linear) motor flow data is fit by a neural response with only a small overshoot (a = 0.25), whereas the (highly nonlinear) visual flow data is fit by a neural response with a large overshoot (a = 3.0). Also note that the flow impulse responses fit to the motor and visual data are quite similar, with FWHM of 5.2 sec (motor) and 6.2 sec (visual).

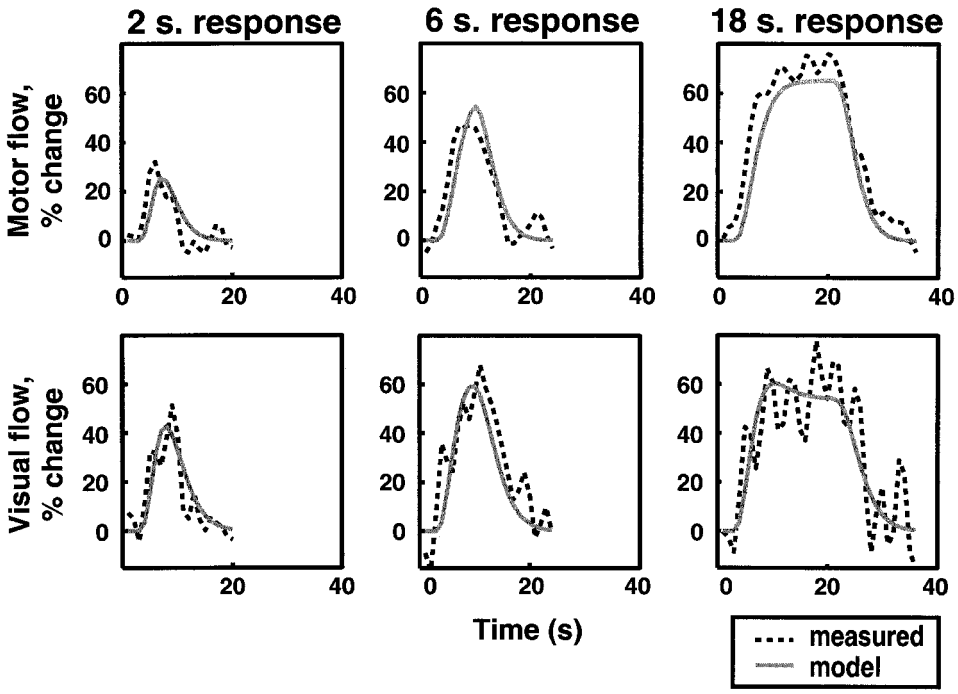

Figure 9.
Flow model fits (solid gray lines) and measured flow responses (dotted black lines). The flow models are generated from the functions shown in Figure 8 by varying the input stimulus duration.

flow response to short stimuli to be disproportionately large, introducing nonlinearities like those found in the visual data set. A smaller overshoot will make the flow response more nearly linear (with linearity for $a = 0$). By varying a we are able to achieve the range of nonlinearities observed in our flow data. The parameters for the flow model fit to the visual data (τ_n and a) are within the range of those found in studies of single-cell recordings in the visual cortex [Albrecht et al., 1984; Maddess et al., 1988; Bonds, 1991].

It is important to note that the differences between the motor and visual responses discussed here are not necessarily fundamental characteristics of the visual and motor cortices, but may reflect differences in the neural response to self-initiated activity compared to passive stimulation. Self-initiated stimulation, such as finger tapping, may involve less neural adaptation than stimuli such as passive viewing.

An interesting subtlety of the motor flow response involves the small dip found in the average 2- and 6-sec responses. One result of such a dip is to counteract the effect of a strong early response in a convolution. In our motor flow data, the small overprediction of the 6-sec response by the 2-sec response is not found when the prediction is extended to 18 sec (i.e., the 2-sec response slightly *underpredicts* the 18-sec response). When the 6-sec prediction is shifted to give an 18-sec prediction, destructive interference of the peak and dip yields a smaller peak response than if no dip were present because the plateau level of a long-duration response depends on just the total area under the impulse response curve. Although our model does not incorporate such a dip, neural responses with a poststimulus dip have been reported in the visual cortex [Smirnakis et al., 1997]. Note that the presence of such a dip does not necessarily indicate nonlinearity. For example, a system responding linearly to rates of change can cause such a response.

The BOLD response

The measured BOLD response is consistently nonlinear in the studied brain areas, with short duration responses overpredicting long duration responses. Clearly, the nonlinearities discussed previously in the flow response are implicitly present in the BOLD signal as well. Additional nonlinearities are introduced in the transition from flow response to BOLD effect [Buxton and Frank, 1997; Friston et al., 1998; Davis et al., 1998; Liu et al., 2000]. Although the transient effects that arise in the transition from flow response to BOLD response make modeling these final steps in Figure 1 complicated [see Buxton et al., 1998; Davis et al., 1998], we can make some basic observations about the relationship between the flow response and the BOLD effect. The BOLD effect saturates at high levels of flow as further increases in flow cause negligible decreases in the concentration of deoxyhemoglobin. This nonlinearity between the flow change and the BOLD signal change is illustrated by the solid gray

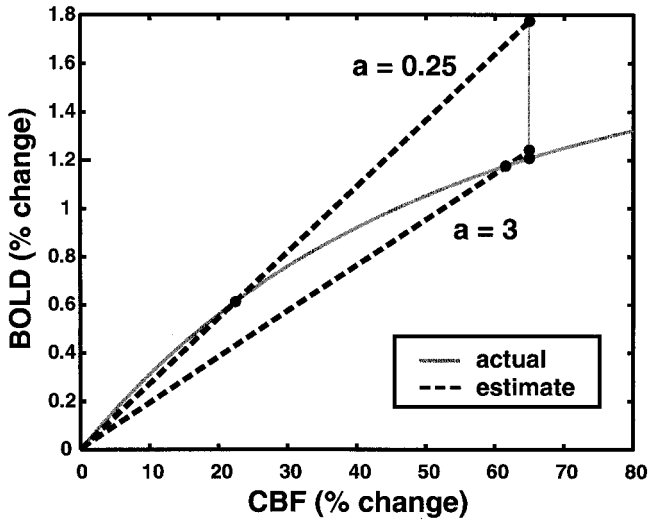

Figure 10.

Illustration of the interaction between neural adaptation nonlinearities and the saturation nonlinearity of the BOLD effect at high levels of flow. The flow signal is assumed to be scaled according to the saturation curve (solid gray line) to produce the BOLD signal. For small adaptation nonlinearities (a nearly linear flow response such as our motor measurements, a = 0.25), the flow change measured for short-duration stimuli is small (here, 22%) compared to the flow response to long-duration stimuli (65%). In this case, the effect of BOLD saturation is strong, and a linear prediction of the long-duration BOLD response based on the short-duration BOLD response (upper dashed black line) will yield a strong overprediction. On the other hand, a strong adaptation nonlinearity (i.e., a strongly nonlinear flow response such as in our visual data, a = 3.0) will cause even short-duration flow responses to nearly reach the plateau value. The saturation nonlinearity has little effect in this case (lower dashed line).

curve in Figure 10, which is similar to the model of Davis et al. [1998]. This relationship causes linear estimates of the BOLD change at high levels of flow to be overpredicted by linear extrapolation of the BOLD change at low levels of flow (see Fig. 10).

It may seem surprising, given multiple sources of nonlinearity in the BOLD signal, that the observed flow nonlinearities are more varied than BOLD nonlinearities, which are fairly consistent across all measurements. A plausible explanation for this phenomenon results from the interaction of nonlinearities in the transition from stimulus to neural activity with nonlinearities introduced between flow and BOLD.

For small neural adaptation effects (small a), the peak flow response to short duration stimuli is small compared to the steady state plateau level. In this case, the step from flow response to BOLD effect scales short duration (low flow) responses by a different amount than it scales long duration (high flow) responses, resulting in the introduction of an overprediction nonlinearity. As the strength of adaptation effects in the neural response increases (larger a), the flow response to short duration stimuli tends to peak closer to the plateau level reached during long stimuli. When peak flow changes are similar, there is little nonlinearity in the scaling step from flow change to BOLD signal change. In other words, when we combine the effects of adaptation and saturation, we find that a weak adaptation nonlinearity will tend to accompany a strong saturation nonlinearity, and vice versa. Consequently, these two effects (adaptation of the neural response and saturation of the BOLD signal) counteract each other to produce a nonlinearity in the BOLD response that is more consistent than nonlinearities in the neural or CBF responses.

As shown by the simulation in Figure 11, this interaction is sufficient to produce model responses that are very similar to our data. When a is small (as in our motor data), representing a nearly linear neural response, the flow response is nearly linear while the BOLD response is distinctly nonlinear. In contrast, when a is large (as in our visual data), representing a strongly nonlinear neural response, the flow response is strongly nonlinear, and little additional nonlinearity is introduced in the final step from flow change to BOLD signal.

Although the above explanation is framed in terms of a flow response that is a linear transform of the nonlinear neural response, our data cannot test this linearity. For example, an alternative hypothesis is that for brief stimuli the CBF response has the same strength independent of stimulus duration [Glover, 1999]. This could be due to a nonlinear step between neural activity and CBF change. However, the same interplay of nonlinearities described above would apply even if the flow is a nonlinear transformation of the neural response. That is, any system giving rise to the observed flow nonlinearities will possess these consistent BOLD nonlinearities when transformed according to the flow-BOLD relationship described by the saturating curve in Figure 10.

In the above discussion, we have ignored the potential effects of the poststimulus undershoot of the BOLD signal [Buxton et al., 1999]. In this transient effect, after the end of stimulation, the BOLD signal drops from the activated state to below the steady-state resting baseline, and then slowly rises back to baseline. If sufficiently long rest periods are not allowed for the response to return to baseline before subsequent stimulation, the poststimulus undershoot is known to cause an apparent shift in baseline [Fransson et al., 1998]. Because the primary focus of this

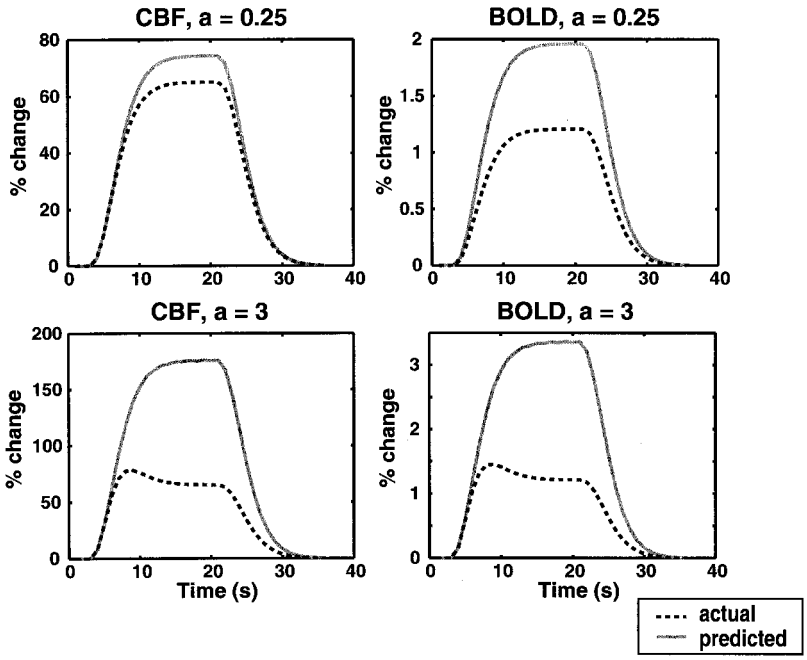

Figure 11.
Simulations of the linearity analyses shown in Figures 4 and 5 using the flow models shown in Figure 8 and the BOLD saturation curve shown in Figure 10. The models were used to generate flow and BOLD responses to 2- and 18-sec stimuli. The curves shown are the "actual" 18-sec response (in black) and the 18-sec response that is predicted from the 2-sec response (in gray). The top figures (labeled a = 0.25) are generated from the flow model fit to the motor flow data. Here, a *weak* nonlinearity in the flow response is transformed into a *strong* BOLD nonlinearity by the BOLD saturation effect. The bottom figures (labeled a = 3.0) are generated from the visual flow model. The *strong* adaptation nonlinearity in the flow response leads to a *weak* nonlinearity in the step from CBF change to BOLD change. Note the similarity of these curves to the corresponding figures marked "2/18 s" in Figures 4 and 5.

work was to investigate the CBF response, which typically does not exhibit strong long-lasting undershoots, we used a relatively short rest period in order to allow more stimulus cycles to be averaged and improve the signal to noise ratio of the flow signal. Our relatively short rest periods (19 sec) are likely to cause baseline shifts in the BOLD signal, meaning that an undershoot would certainly affect the measured nonlinearities of the BOLD signal. However, the form of the nonlinearity (overprediction or underprediction) depends on the details of how the undershoot duration and amplitude scale with stimulus duration. For example, a poststimulus undershoot that scales in proportion to the primary (positive) BOLD signal change would have a minor effect on the apparent linearity of the BOLD signal measured with short off-periods. Detailed studies using sufficiently long off-periods (at least 60 sec) will be required to determine the effect of poststimulus undershoot on the linearity of the BOLD response.

Another important consideration in the nonlinear characteristics described here is the possible effects of voxel selection criteria. Voxels were selected based on a requirement of significant linear flow signal. While this selection criterion does not affect our ability to detect nonlinearity, it does mean that voxels with extreme nonlinear flow responses (such as a spiked response at the beginning of a stimulus that immediately returns to baseline for the remainder of the stimulus) may have been excluded from our analysis. Although no data currently supports this type of response, this possibility warrants further study.

CONCLUSIONS

The CBF response to brief stimuli in the primary motor and visual cortices is in general a nonlinear function of stimulus duration, but the degree of nonlinearity varies strongly across the two regions. This finding reflects the variability of the flow response but not necessarily a characteristic difference between these two cerebral regions. In contrast, the BOLD response is consistently nonlinear, with the response to shorter stimuli overpredicting the response to longer stimuli. This pattern is consistent with a model in which the steps from stimulus to neural activity and from CBF change to BOLD signal change are described by known nonlinearities, but the step from neural activity to CBF change is linear. However, further experiments directly comparing neural activity with CBF changes will be required to establish this linearity. Such a linear relationship would imply that arterial spin labeling measurements of CBF change are a more faithful reflection of neural activity than the BOLD response.

ACKNOWLEDGMENTS

We thank Dr. Geoffrey Boynton for discussions on neural response modeling, and Drs. Steven Conolly

and John Pauly for discussions on signal processing. Dr. Frank is supported by VA Merit Review SA 321.

REFERENCES

Albrecht D, Farrar S, Hamilton D (1984): Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex. J Physiol 347:713–739.

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

Bandettini P, Wong E, Hinks R, Tikofsky R, Hyde J (1992): Time course EPI of human brain function during task activation. Magn Reson Med 25:390–397.

Bonds A (1991): Temporal dynamics of contrast gain in single cells of the cat striate cortex. Vis Neurosci 6:239–255.

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

Buchel C, Holmes A, Rees G, Friston K (1998): Characterizing stimulus-response functions using non-linear regressors in parametric fMRI experiments. Neuroimage 8:140–148.

Buxton R, Frank L (1997): A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. J Cereb Blood Flow Metab 17:64–72.

Buxton R, Wong E, Frank L (1997): A comparison of perfusion and BOLD changes during brain activation. Proc ISMRM, p 153.

Buxton R, Wong E, Frank L (1998): Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn Reson Med 39:855–864.

Buxton R, Wong E, Frank L (1999): The post-stimulus undershoot of the functional MRI signal. In: Moonen C, Bandettini P, editors. Functional MRI. Berlin: Springer-Verlag. p 253–262.

Cohen J, MacWhinney B, Flatt M, Provost J (1993): PsyScope: a new graphic alternative environment for designing psychology experiments. Behav Res Meth Instr Comp 25:257–271.

Cox R (1996): AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comp Biomed Res 29:162–173.

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

Davis T, Kwong K, Weisskoff R, Rosen B (1998): Calibrated functional MRI: mapping the dynamics of oxidative metabolism. Proc Natl Acad Sci USA 95:1834–1839.

Duyn J, Tan C, van der Veen J, van Gelderen P, Frank J, Ye F, Yongbi M (2000): Perfusion-weighted single-trial fMRI. Proc ISMRM, p 55.

Fox P, Raichle M (1986): Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. Proc Natl Acad Sci USA 83:1140–1144.

Fox P, Raichle M, Mintun M, Dence C (1988): Nonoxidative glucose consumption during focal physiologic neural activity. Science 241:462–464.

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

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

Glover G (1999): Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage 9:416–429.

Hoge R, Atkinson J, Gill B, Crelier G, Marrett S, Pike G (1999): Stimulus-dependent BOLD and perfusion dynamics in human V1. Neuroimage 9:573–585.

Kim S-G (1995): Quantification of relative cerebral blood flow change by flow sensitive alternating inversion recovery (FAIR) technique: application to functional mapping. Magn Reson Med 34:293–301.

Kwong K, Belliveau J, Chesler D, Goldberg I, Weisskoff R, BP, P, Kennedy D, Hoppel B, Cohen M, Turner R, Cheng H-M, Brady T, Rosen B (1992): Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci USA 89:5675–5679.

Lange N, Zeger S (1997): Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging. Appl Statistics 46:1–29.

Liu H-L, Gao J-H (1999): Perfusion-based event-related functional MRI. Magn Reson Med 42:1011–1013.

Liu T, Luh W-M, Wong E, Bandettini P, Obata T, Frank L, Buxton R (2000): On the nonlinear relation between BOLD and CBF. Proc ISMRM, p 948.

Maddess T, McCourt M, Blakeslee B, Cunningham R (1988): Factors governing the adaptation of cells in area-17 of the cat visual cortex. Biol Cybern 59:229–236.

Miller K, Buxton R, Wong E, Frank L (1999): The linearity of the cerebral blood flow response to brief motor tasks. Proc ISMRM, p 381.

Miller K, Luh W-M, Liu T, Martinez A, Obata T, Wong E, Frank L, Buxton R (2000): Characterizing the dynamic perfusion response to stimuli of short duration. Proc ISMRM, p 500.

Ogawa S, Tank D, Menon R, Ellermann J, Kim S-G, Merkle H, Ugurbil K (1992): Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proc Natl Acad Sci USA 89:5951–5955.

Rees G, Howseman A, Josephs O, Frith C, Friston K, Frackowiak R, Turner R (1997): Characterizing the relationship between BOLD contrast and regional cerebral blood flow measurements by varying the stimulus presentation rate. Neuroimage 6:270–278.

Rieke F, Warland D, Ruyter van Steveninck R, Bialek W (1997): Spikes: exploring the neural code. Cambridge, MA: MIT Press.

Robson M, Dorosz J, Gore J (1998): Measurements of the temporal fMRI response of the human auditory cortex to trains of tones. Neuroimage 7:185–198.

Smirnakis S, Berry M, Warland D, Bialek W, Meister M (1997): Adaptation of retinal processing to image contrast and spatial scale. Nature 386:69–73.

Vazquez A, Noll D (1998): Nonlinear aspects of the BOLD response in functional MRI. Neuroimage 7:108–118.

Wong E, Bandettini P, Hyde J (1992): Echo-planar imaging of the human brain using a three axis local gradient coil. Proc SMRM, p 105.

Wong E, Buxton R, Frank L (1997): Implementation of quantitative perfusion imaging techniques for functional brain mapping using pulsed arterial spin labeling. NMR Biomed 10:237–249.

Wong E, Buxton R, Frank L (1998): Quantitative imaging of perfusion using a single subtraction (QUIPSS and QUIPSS II). Magn Reson Med 39:702–708.

Yang Y, Engelien W, Pan H, Xu S, Silbersweig D, Stern E (2000): A CBF-based event-related brain activation paradigm: characterization of impulse-response function and comparison to BOLD. Proc ISMRM, p 237.

Ye F, Smith A, Yang Y, Duyn J, Mattay V, Ruttimann U, Frank J, Weinberger D, McLaughlin A (1997): Quantification of regional cerebral blood flow increases during motor activation: a steady-state arterial spin tagging study. Neuroimage 6:104–112.