COMMUNICATIONS

Dynamics of Blood Flow and Oxygenation Changes During Brain Activation: The Balloon Model

Richard B. Buxton, Eric C. Wong, Lawrence R. Frank

A biomechanical model is presented for the dynamic changes in deoxyhemoglobin content during brain activation. The model incorporates the conflicting effects of dynamic changes in both blood oxygenation and blood volume. Calculations based on the model show pronounced transients in the deoxyhemoglobin content and the blood oxygenation level dependent (BOLD) signal measured with functional MRI, including initial dips and overshoots and a prolonged post-stimulus undershoot of the BOLD signal. Furthermore, these transient effects can occur in the presence of tight coupling of cerebral blood flow and oxygen metabolism throughout the activation period. An initial test of the model against experimental measurements of flow and BOLD changes during a finger-tapping task showed good agreement.

Key words: functional magnetic resonance imaging; cerebral blood volume; cerebral blood flow; cerebral oxygen metabolism.

INTRODUCTION

Despite the widespread use of functional neuroimaging techniques, the physiological changes in the brain that accompany neural activation are still poorly understood. During brain activation, a modest increase in the cerebral metabolic rate of oxygen ($CMRO_2$) is accompanied by a much larger increase in local blood flow (1, 2). Because of this imbalance, local capillary and venous blood are more oxygenated during activation. The large increase in flow is the basis for mapping brain activation patterns with positron emission tomography (PET), and the decrease in local deoxyhemoglobin concentration is the basis for functional magnetic resonance imaging (fMRI) exploiting the blood oxygenation level dependent (BOLD) effect (3–7). Recently, we proposed a possible explanation for the large flow increase as a result of tight coupling of flow and oxygen metabolism in the presence of limited oxygen delivery (8). This oxygen limitation model is based on the assumptions that essentially all of the oxygen that leaves the capillary is metabolized and that blood flow increases are accomplished by increased capillary blood velocity rather than capillary recruitment. Under these circumstances, increased blood flow leads to reduced oxygen extraction due to the decreased capillary transit time. The rate of delivery of oxygen, which is proportional to the product of flow and oxygen extraction fraction, therefore increases much less than the flow itself. In the context of this model, a large increase in flow is required to support a small increase in oxygen metabolism.

The oxygen limitation model provides an interpretation of the mismatch between flow and oxygen metabolism changes during brain activation as evidence for tight coupling, rather than uncoupling, of flow and metabolism. However, detailed studies of the time course of deoxy- and oxyhemoglobin changes during activation have revealed a complex pattern that may contain evidence of a transient uncoupling of flow and metabolism. In optical studies in a cat model, Malonek and Grinvald (9) found that after a brief (4-s) visual stimulation, the total deoxyhemoglobin increased for the first 3 s, followed after a few seconds by the larger decrease associated with the BOLD effect. This "fast response" was interpreted as an initial increase in oxygen extraction, before the large flow increase.

In addition, there is now a substantial amount of experimental data measuring the time course of the BOLD effect. A square wave pattern of stimulus presentation typically leads to a BOLD response that is delayed by 2 to 3 s followed by a ramp of 6 to 10 s to a plateau value, followed by a return to baseline after the stimulus with a similar ramp (10). However, the BOLD signal time course during brain activation also has been reported to exhibit several transient features at the onset and end of the stimulus: an initial dip (11–13), corresponding to the initial increase in local deoxyhemoglobin observed with optical techniques (9); a post-stimulus undershoot not evident in the in-flow signal (14); and an initial overshoot followed by a much weaker plateau and a strong post-stimulus undershoot (15, 16). A possible explanation for these BOLD effects is a transient uncoupling of flow and oxygen metabolism. The initial dip has been interpreted as evidence for an initial increase in oxygen extraction before the flow increase (9). The post-stimulus undershoot has been modeled as an elevated oxygen extraction, after flow has returned to baseline, required to replenish depleted tissue oxygen stores (17). The initial overshoot, weak plateau, and post-stimulus undershoot have been interpreted as evidence for transient uncoupling, recoupling, and uncoupling again of flow and oxygen metabolism (15, 16).

MRM 39:855–864 (1998)

From the Departments of Radiology (R.B.B., E.C.W., L.R.F.) and Psychiatry (E.C.W.), University of California at San Diego, San Diego, California.

Address correspondence to: Richard B. Buxton, Ph.D., Associate Professor of Radiology, UCSD Medical Center, 200 West Arbor Drive, San Diego, CA 92103-8756.

Received July 3, 1997; revised November 28, 1997; accepted February 9, 1998.

0740-3194/98 $3.00

Thus, the observed transients in signals dependent on blood oxygenation may be evidence for uncoupling of flow and metabolism. However, recent experimental work suggests another explanation for these phenomena. Mandeville *et al.* (18, 19) found in animal studies with an intravascular marker that the dynamics of blood volume change do not match the dynamics of the flow change. Although the deoxyhemoglobin concentration in blood depends on flow and oxygen extraction, the total amount of deoxyhemoglobin depends strongly on blood volume, as well. For this reason, a quantitative understanding of the combined effects of changes in flow, oxygen extraction, and blood volume is required to clarify the significance of the observed transients in the BOLD signal. In this article, we describe a simple biomechanical model for the blood volume changes during brain activation that is consistent with tight coupling of flow and oxygen metabolism throughout the activation and yet produces transients similar to those observed experimentally. A preliminary version of this work has been presented in abstract form (20).

THE BOLD SIGNAL

The mathematical model presented here describes the changes in physiological variables during brain activation. To connect the model with experimental fMRI, we first need a quantitative model for BOLD signal changes as a function of blood susceptibility and volume. This relationship has been extensively explored in recent years using experimental data (21), numerical Monte Carlo simulations (21–23), and analytical calculations (24). Although the signal dependence is intrinsically a nonlinear function of susceptibility and blood volume due to the effects of diffusion, the situation is relatively simple for the case of gradient echo (GRE) signals affected by postcapillary blood vessels. In this initial analysis, we make the simplifying assumption that the GRE signal changes are primarily due to small postcapillary venous vessels and neglect the contribution of capillaries. In this case, the role of diffusion is minor because of the larger size of the venous vessels, and the extravascular signal changes essentially depend just on the change in the total amount of deoxyhemoglobin in the tissue voxel (23). In addition, Boxerman *et al.* (25) have recently investigated the role of intravascular signal changes in BOLD experiments and conclude that at 1.5 T, these changes contribute more than half of the net observed signal change. We therefore include both contributions in our model for the BOLD signal.

As a starting point, we consider the total BOLD signal to be a volume-weighted sum of the extravascular (S_e) and intravascular (S_i) signal:

$$S = (1 - V)S_e + VS_i \qquad [1]$$

where V is the blood volume fraction. Then, for small signal changes ΔS:

$$\Delta S = (1 - V_0)\Delta S_e - \Delta V\, S_e + V_0 \Delta S_i + \Delta V\, S_i \qquad [2]$$

where V_0 is the resting blood volume fraction. The signal then depends on how the intrinsic signals change when the deoxyhemoglobin content and the volume change. In theoretical development, we will use the notation that capital letters refer to a specific dimensional value of a parameter, whereas lower case letters refer to the value during activation normalized to the value at rest. We write the normalized total deoxyhemoglobin content as $q = Q/Q_0$ and the normalized volume as $v = V/V_0$. For the change in the extravascular component, we use the numerical results of Ogawa *et al.* (23), showing that for the small venous vessels, the transverse relaxation rate (R_2^*) is proportional to the product of the blood deoxyhemoglobin concentration and blood volume, so ΔS_e depends only on the total deoxyhemoglobin *q*. For the intravascular component (and for the intrinsic ratio S_i/S_e), we use the numerical results of Boxerman *et al.* (25). From their calculations (their Fig. 2), ΔS_i can be reasonably approximated as a linear function of the deoxyhemoglobin concentration in blood for O_2 extraction fractions between 20% and 50% (the range of interest for brain). Combining these numerical results with Eq. [2] leads to the following form for the BOLD signal (see Appendix for details):

$$\frac{\Delta S}{S} = V_0\left[k_1(1 - q) + k_2\left(1 - \frac{q}{v}\right) + k_3(1 - v)\right] \qquad [3]$$

The first term describes the intrinsic extravascular signal, the second term describes the intravascular signal, and the third term describes the effect of changing the balance of the sum in Eq. [1]. The parameters k_1, k_2, and k_3 are dimensionless and can be estimated from the earlier work. For $B_0 = 1.5$ T and TE = 40 ms, we estimate that $k_1 \cong 7\,E_0$ based on the numerical studies of Ogawa *et al.* (23), where E_0 is the resting oxygen extraction fraction. Based on the results of Boxerman *et al.* (25), we estimate $k_2 \cong 2$ and $k_3 \cong 2\,E_0 - 0.2$. In Eq. [3], q is the normalized total voxel content of deoxyhemoglobin (i.e., $q = 1$ at rest), v is the normalized venous blood volume ($v = 1$ at rest), and V_0 is the actual venous blood volume fraction (e.g., 1–4%).

BALLOON MODEL

To model the transient aspects of the BOLD signal, we assume again that there is no capillary recruitment, in keeping with the oxygen limitation model above, so that blood volume changes occur primarily in the venous compartment. (The arteriolar dilitation that produces a flow increase is assumed to be a negligible change in blood volume.) The vascular bed within a small volume of tissue is then modeled as an expandable venous compartment (a balloon) that is fed by the output of the capillary bed. The volume flow rate (ml/s) into the tissue, $F_{in}(t)$, is an assumed function of time that drives the system. The volume flow out of the system, $F_{out}(t)$, is assumed to depend primarily on the pressure in the venous compartment. For example, we could write that $F_{out} = (P - P_{mixed})/R$, where P is the pressure in the venous compartment, P_{mixed} is the mixed venous pressure downstream from the tissue element, and R is the resistance of the vessels after the balloon. Then, the physical picture of activation is that after the arteriolar resis-

tance is decreased, producing an increase in F_{in}, the venous balloon swells, and the pressure increases until F_{out} matches F_{in} (a new steady state). The amount of swelling that occurs will depend on the biomechanical properties of the vessel as reflected in the pressure/volume curve of the venous balloon. However, rather than introduce pressures explicitly into the equations (and thus introduce more parameters), we simply assume that F_{out} is a function of the venous volume, V. The rate of change of the volume of the balloon is the difference between F_{in} and F_{out}. We can then consider the total deoxyhemoglobin Q(t) in the tissue element. For simplicity, we neglect the capillary contribution and assume that all of the deoxyhemoglobin is in the venous compartment. The rate of entry of deoxyhemoglobin into the venous compartment is F_{in} E C_a, where E is the net extraction of O_2 from the blood as it passes through the capillary bed and C_a is the arterial O_2 concentration (assumed to be due to a fully oxygenated hemoglobin concentration). If we then treat the balloon as a well-mixed compartment, the clearance rate of deoxyhemoglobin from the tissue is $F_{out} \times$ the average venous concentration, Q(t)/V(t). The coupled equations for Q(t) and V(t) are then

$$\frac{dQ}{dt} = F_{in}(t)EC_a - F_{out}(V)\frac{Q(t)}{V(t)}$$
$$\frac{dV}{dt} = F_{in}(t) - F_{out}(V) \qquad [4]$$

By scaling each of these variables with their value at rest (t = 0), these equations can be written as in the previous section in terms of the dimensionless variables $q(t) = Q(t)/Q_0$, $v(t) = V(t)/V_0$, $f_{in}(t) = F_{in}(t)/F_0$, and $f_{out}(v) = F_{out}(V)/F_0$:

$$\frac{dq}{dt} = \frac{1}{\tau_0}\left[f_{in}(t)\frac{E(t)}{E_0} - f_{out}(v)\frac{q(t)}{v(t)}\right]$$
$$\frac{dv}{dt} = \frac{1}{\tau_0}[f_{in}(t) - f_{out}(v)] \qquad [5]$$

where Q_0 is the resting deoxyhemoglobin content, V_0 is the resting volume, F_0 is the resting flow, $\tau_0 = V_0/F_0$ is the mean transit time through the venous compartment at rest, E_0 is the resting net extraction of O_2 by the capillary bed, and $q(0) = v(0) = f_{in}(0) = f_{out}(v(0)) = 1$. Note that τ_0 simply sets the time scale for changes and that the only other parameter that appears explicitly is E_0. However, there are two functions that remain to be specified: E(t) and $f_{out}(v)$. At this point, we can include the oxygen limitation model by requiring that the net extraction is equal to the unidirectional extraction. That is, we assume that $CMRO_2$ increases as much as possible within the constraints set by limited oxygen delivery. Under these circumstances, we argued previously that a nonlinear expression for E(f) is a reasonable approximation for a wide range of transport conditions (8):

$$E(f) = 1 - (1 - E_0)^{1/f} \qquad [6]$$

In Eq. [5], E(t) is then replaced with $E(f_{in}(t))$ from Eq. [6]. Variable temporal patterns of response will then be created for different functional forms of $f_{out}(v)$, which is equivalent to different pressure/volume curves for the venous balloon.

In constructing $f_{out}(v)$, we need to distinguish between the steady-state relationship of blood flow and volume and the transient relation $f_{out}(v)$ that controls the transition from one steady state to another. In early work using altered CO_2 to change blood flow, Grubb *et al.* (26) found that the steady-state total blood volume could be described empirically by a power law relationship, $v = f^{\alpha}$, where α was found to be about 0.4 from their data. The simplest assumption would be that $f_{out}(v)$ follows this same empirical relation. However, experimental measurements by Mandeville *et al.* (18, 19), using a vascular contrast agent, suggested that although the steady-state relationship was in reasonable agreement with Grubb *et al.* (26), the transition periods were not. Therefore, for the numerical studies described here, we assumed that the steady-state plateau level of volume change is given by a power law with exponent α, but the time course during the transition between steady-state levels is different. In practical terms for the calculations, this means that the end points of the curve $f_{out}(v)$ are specified by the power law relation, but the shape of the curve between those end points is variable. For the calculations, we modeled $f_{out}(v)$ as a sum of a linear component and a power law.

EXPERIMENTAL METHODS

As an initial test of the balloon model, we measured the time course of cerebral blood flow and BOLD signal changes during a simple finger-tapping task. Measurements were made with a new pulsed arterial spin labeling technique recently developed in our laboratory, called Quantitative Imaging of Perfusion with a Single Subtraction, version II (QUIPPS II) (27, 28). Arterial spin labeling involves tagging the arterial blood with an inversion pulse before entering the slice of interest, a delay to allow blood to enter the slice, followed by collection of a rapid echo planar image, creating the tagged image. The pulse sequence is then repeated without the initial inversion to create a control image. The difference image, control minus tagged, is then directly proportional to the amount of blood delivered during the delay period and thus is directly proportional to local perfusion. QUIPSS II is specifically designed to deal with the confounding effects of nonuniform transit delays from the tagging region to different brain regions, which can severely compromise perfusion estimates based on measurements with a single inversion delay. In QUIPSS II, at time TI_1 after the inversion pulse, a saturation pulse is applied to the tagging region, destroying that portion of the tagged blood that remains in the tagging region and thus forcing the time width of the tag to equal TI_1. The image is then acquired at a later time, TI_2, to allow the tagged bolus of blood to reach all of the voxels in the imaged slice. Provided that the delay $TI_2 - TI_1$ is longer than the largest transit delay, then all of the tagged blood destined for each voxel will have arrived.

For the current studies, we used an off-resonance control RF pulse, identical to the inversion pulse, but with no slice selective field gradients (PICORE) (28), to produce similar off-resonance effects on the imaged slice without tagging arterial blood. We used a GE 1.5-T SIGNA clinical MRI scanner (General Electric Corp., Milwaukee, WI) fitted with highperformance local head gradient and RF coils of our own design and construction (29). Pulse sequence parameters for these gradient echo acquisitions were TI_1 = 600 ms, TI_2 = 1200 ms, TE = 30 ms, TR = 2.011 s, 64 × 64 echo-planar imaging acquisition with in-plane resolution 3.75 × 3.75 mm and slice thickness 8 mm. Repetition times on the order of 2 s are required to allow replenishment of fully relaxed blood. To increase our temporal resolution, the TR was chosen to be slightly asynchronous with the stimulus period so that when multiple cycles are collapsed to form an average single cycle time course, later cycles provide measurements in-between the measurements of the first cycle, producing a time resolution of approximately 1 s.

Six healthy volunteers were asked to perform bilateral sequential finger tapping for 1 min, followed by 2 min of rest. This 3-min cycle was repeated eight times, while 750 images of a single axial slice passing through the motor areas were collected alternating between tag and control images. In the raw data time course, the flow signal is encoded in the image-to-image difference signal, and the BOLD signal is encoded in the slow variations of the average signal after the stimulus. The images are alternately positively and negatively flow-weighted by the arterial tag. To remove the flow weighting from the BOLD signal, each image is added to the average of its two neighbors in time. To minimize the BOLD weighting in the flow signal, each image is subtracted from the average of its neighbors. Each time course is then collapsed appropriately over cycles, to produce an average time course for one cycle for each pixel. For each subject, two 8 × 8 block of pixels covering the left and right motor areas were selected for further analysis. From these 128 pixels per subject, all pixels showing a change with activation in the flow signal of >40% were selected, and the flow and BOLD time courses for these pixels were averaged across all subjects (a total of 153 pixels) to produce high signal-to-noise dynamic curves. In the reported data, the amplitude of the flow signal is plotted as a percentage of the total mean brain signal. Absolute quantitation in units of cerebral blood flow (CBF) requires a global scaling factor that depends on the proton density and relaxation times of blood. This calibration factor was not measured, but we estimate for assumed relaxation parameters that a 1% flow signal corresponds to CBF of 50 to 60 ml/min-ml of tissue.

RESULTS

The curves in the figures illustrate the variety of time courses that can be produced by the balloon model. These curves were calculated by numerical integration of Eq. [5] using the Mathematica software package (Wolfram Research). For all of these calculations, τ_0 = 2 s and E_0 = 0.4, and the driving function of the system $f_{in}(t)$ was assumed to be a trapezoidal function with a rise time of 4 to 6 s and variable duration. These calculations describe the response of the system to a given inflow curve. That inflow curve, in turn, is presumed to be a response to a particular neural stimulus. However, the relationship between stimulus and flow response is poorly understood, and in our calculations, we do not attempt to model it. In this section, we will refer to the "duration" of

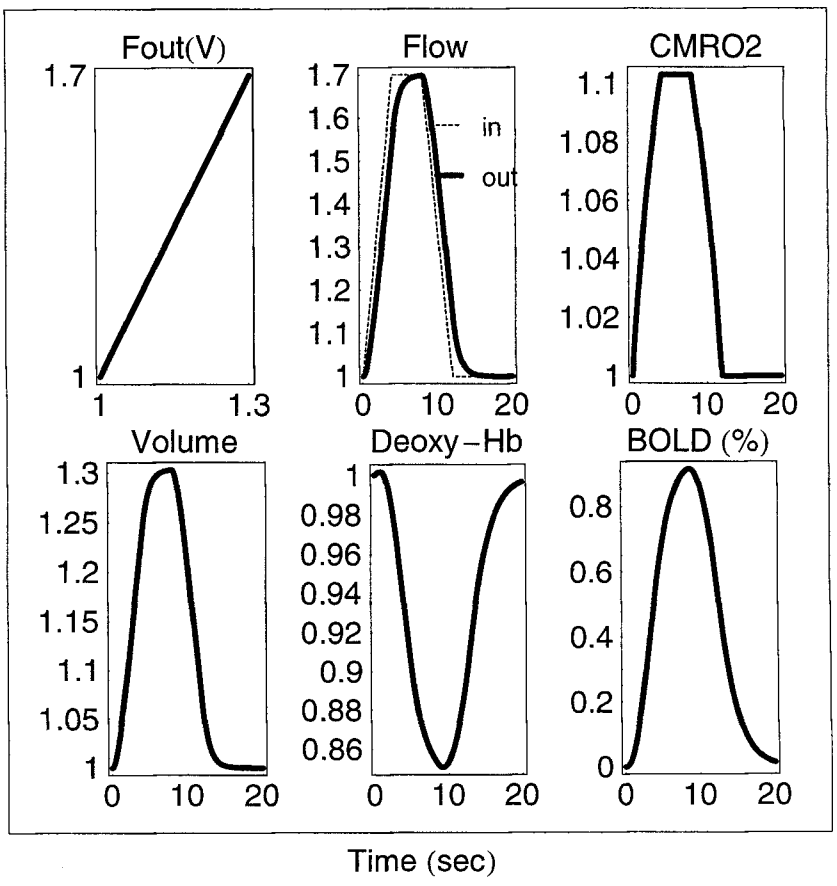

FIG. 1. Balloon model curves for a linear form of $f_{out}(v)$, showing a simple pattern of response. The scale for the BOLD plot is percent signal change, and this amplitude scales directly with the blood volume V_0. Additional model parameters in these calculations were E_0 = 0.4, V_0 = 0.01, and α = 0.5.

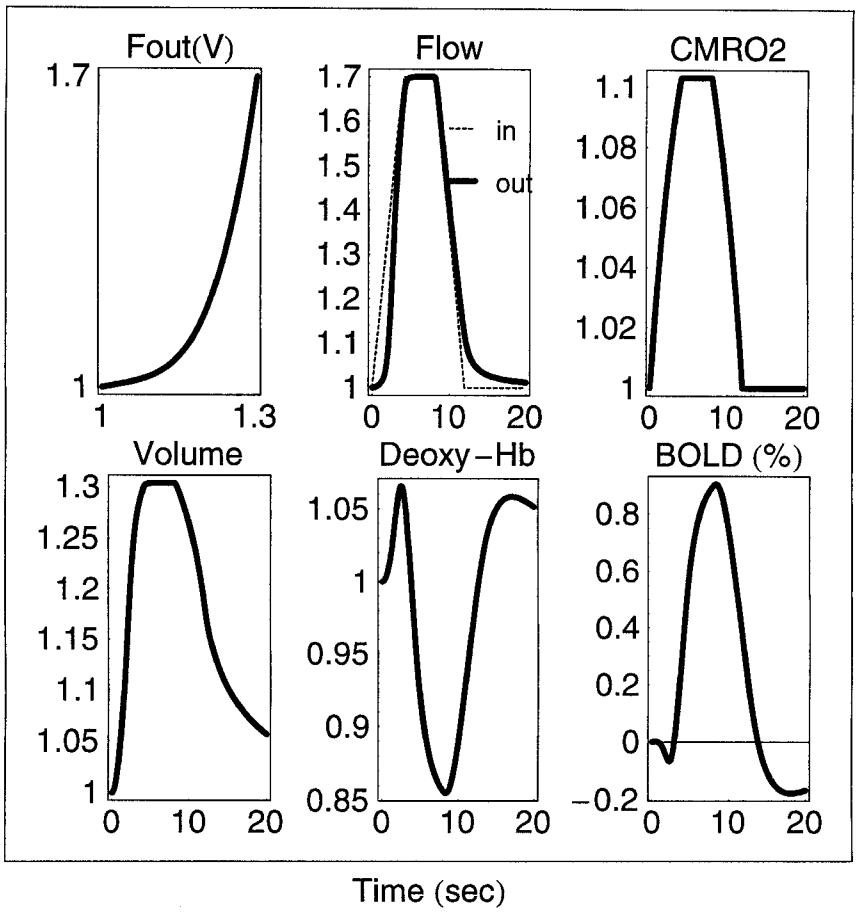

FIG. 2. Balloon model curves for a nonlinear form of $f_{out}(v)$, which produces a more complex pattern of transients. Outflow lags behind inflow, producing an initial fast rise in blood volume and a slow return to baseline. The total deoxyhemoglobin content briefly increases (the fast response), then drops sharply, and finally overshoots the resting level after the flow has returned to the resting level. The BOLD response is qualitatively the inverse of the total deoxyhemolgobin response, showing a weak initial dip, followed by a signal increase, and ending with a prolonged post-stimulus undershoot. Note, however, that the initial dip is weaker than the post-stimulus undershoot despite the fact that the total deoxyhemoglobin change is slightly larger during the fast response. Inflow and $CMRO_2$ do not show any of the transients seen in the deoxyhemoglobin or BOLD signals. Additional model parameters in these calculations were $E_0 = 0.4$, $V_0 = 0.01$, and $\alpha = 0.5$.

an activation as the time from the beginning of the up ramp of $f_{in}(t)$ to the beginning of the down ramp. In each figure, the time courses for different dynamical quantities are shown in separate labeled panels.

Figure 1 shows the calculated response to a short, 8-s duration stimulus with a maximum flow change of 70% when $f_{out}(v)$ is a linear function. Outflow lags slightly behind inflow, but all of the curves show a simple pattern. In Fig. 2, the only change was to make $f_{out}(v)$ a strongly nonlinear function, which produces much more complex transient behavior. There is an initial sharp increase in total deoxyhemoglobin (the fast response), followed by a sharp decrease, and then an overshoot above baseline. The BOLD signal shows the corresponding inverse pattern, including an initial dip and a prolonged post-stimulus undershoot. Note, however, that the initial dip is much weaker than the post-stimulus undershoot, despite the fact that the corresponding changes in total deoxyhemoglobin are similar. This unexpected finding is discussed in more detail below. Despite these pronounced transient features, $CMRO_2$ is tightly coupled to $f_{in}(t)$ throughout the activation period.

Figure 3 shows a comparison of the experimental data with balloon model calculations. The measured flow signal indicates an approximately 70% increase in flow and shows a roughly trapezoidal shape. The BOLD data show a similar shape, but with a pronounced post-stimulus undershoot that is not present in the flow signal. The observed discrepancy between the flow and BOLD signals confirms earlier studies of the post-stimulus undershoot (14). The balloon model produces curves that are very similar to the experimental data for the model parameters and $f_{out}(v)$ curve shown in Fig. 2. The model BOLD curve also shows a small initial dip, but the temporal resolution and signal-to-noise ratio of the experimental data are not sufficient to detect such a small signal change.

Figure 4 shows more specifically how the post-stimulus undershoot comes about in the context of the balloon model. The undershoot occurs when the volume returns to baseline more slowly than flow, producing an elevated total deoxyhemoglobin just due to an increased volume of blood, even though the blood oxygenation has returned to baseline. The rate at which the excess blood volume clears from the tissue depends on the difference between outflow and inflow. Thus, a prolonged post-stimulus undershoot occurs when f_{out} returns nearly to baseline while the volume is still high. More specifically, the duration of the undershoot depends inversely on the initial slope of $f_{out}(v)$. Figure 4 shows that a very subtle change in the initial slope has a strong effect on the undershoot duration.

Figure 5 illustrates the dynamics of the fast response in the context of the balloon model. To emphasize these effects, a short inflow duration of 4 s was used (i.e., a 4-s up ramp immediately followed by a 4-s down ramp). The model parameters and form of $f_{out}(v)$ were the same as in Figs. 2 and 3. The curves show an initial increase in deoxyhemoglobin, peaking around 2.5 s after the beginning of the flow up-ramp, followed by a decrease in deoxyhemoglobin that reaches a minimum at about 6 s.

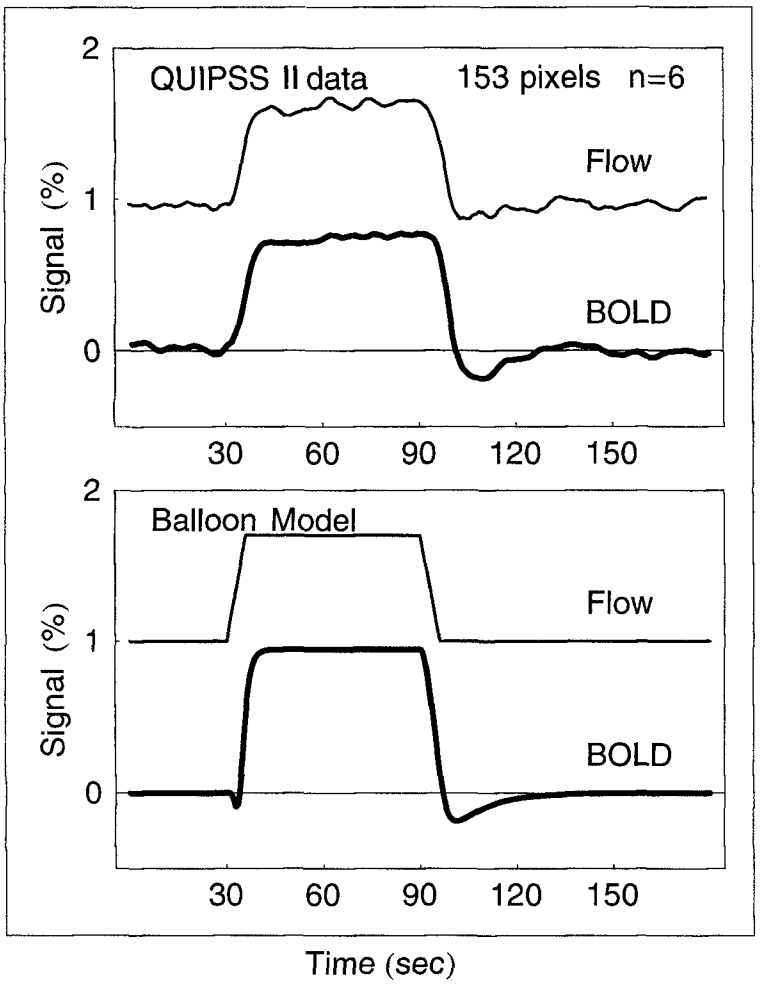

FIG. 3. Experimental data measured with an arterial spin labeling method (QUIPSS II) that allows simultaneous measurement of flow and BOLD signals (top), compared with the balloon model curves (bottom) for the same model parameters used in Fig. 2.

The precise quantitative features of this curve, such as the relative amplitudes of the positive and negative deoxyhemoglobin changes, depend on the specific model parameters and the assumed inflow function. Nevertheless, this temporal pattern is similar to that found by Malonek and Grinvald (9) in optical studies in a cat model with visual stimulation. They found an initial deoxyhemoglobin peak approximately 2 s after the beginning of the visual stimulus, which then decreased to a minimum at about 6 s. In Fig. 5, the initial rise in total deoxyhemoglobin is primarily due to increased blood volume rather than altered blood oxygenation. As flow continues to increase, the blood oxygenation increases and eventually dominates over the volume change, reducing the total deoxyhemoglobin.

Figure 5 also provides an explanation for why the initial dip in the BOLD signal is weaker than the post-stimulus undershoot, despite the fact that the deoxyhemoglobin changes are similar. The lower three panels show separate plots of the intrinsic signal contribution from the intra- and extravascular BOLD signals and the net BOLD signal, which is a weighted average of the other two. The fast response is dominated by a volume change, which increases total deoxyhemoglobin and creates a dip in the extravascular BOLD signal. However, the intravascular signal shows a smooth increase due to the steadily increasing blood oxygenation and is contributing a moderately strong positive signal change at the time of the dip when the extravascular signal change is negative. Because these two signals contribute roughly equally to the net signal at 1.5 T, the dip is nearly cancelled in the net BOLD signal. At higher fields, the BOLD signal is likely dominated by the extravascular BOLD effect, so the dip may be more pronounced. This may partly explain, in addition to signal-to-noise considerations, why the initial dip is rarely seen at lower fields. In contrast, during the post-stimulus undershoot, the intrinsic intravascular signal has returned to baseline because blood oxygenation has returned to baseline, and therefore does not conflict with the undershoot of the extravascular signal.

Figure 6 illustrates the effects of rapidly cycling the stimulus presentations. In this example, a 10-s duration inflow pattern is cycled with a period of 20 s. Thus, the repeated stimuli occur during the post-stimulus undershoot of the previous stimulus. As a consequence, the blood volume never returns to baseline, and the apparent baseline of the BOLD signal is lowered. In addition, the curve of total deoxyhemoglobin is altered in a way that also looks like a shift of the baseline.

In Figure 7, we extended the simple forms of $f_{out}(v)$ used in the other figures to try to produce time courses similar to those found by Frahm and coworkers (15, 16). They reported an initial overshoot in the BOLD signal, followed by a slow decrease to a modest (or weak) plateau level, followed by a pronounced post-stimulus undershoot. The calculations in Fig. 7 differ in two ways from the previous figures: (1) the blood volume change is larger ($\alpha = 0.7$), and (2) there is hysteresis in the curve of $f_{out}(v)$. Given the viscoelastic nature of blood vessels, it is likely that the pressure/volume curve during inflation may be different from that during deflation due to stress relaxation effects. In Fig. 7, $f_{out}(v)$ during inflation increases rapidly with increasing volume, but shows a curve similar to that used in the previous figures at the end of the stimulus. The behavior of the curves at the beginning of the inflow increase is opposite to that in Fig.

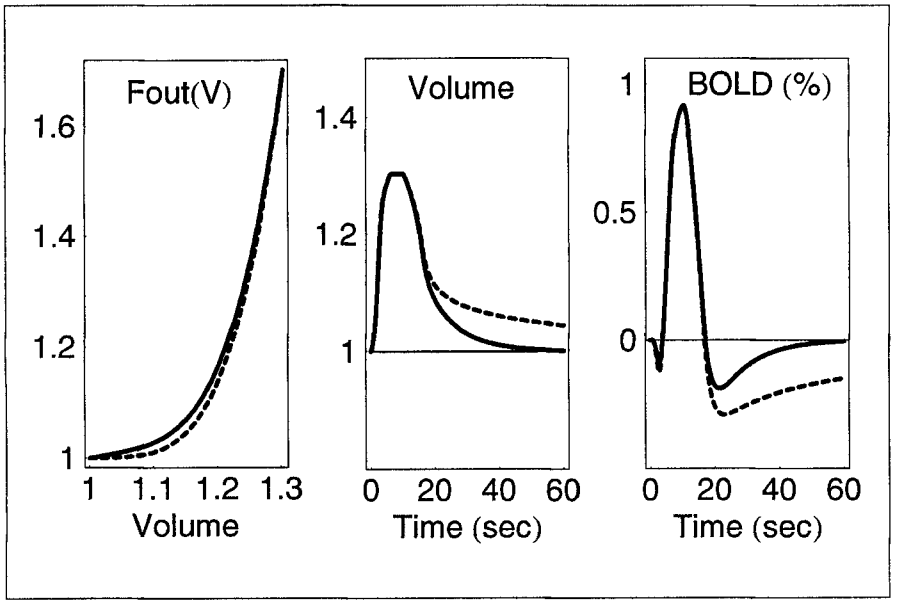

FIG. 4. Curves for a 10-s duration of inflow, illustrating the post-stimulus undershoot. The duration of the undershoot depends inversely on the initial slope of the curve $f_{out}(v)$ (left panel). The slightly steeper slope (solid line) causes the volume change (middle panel) and the BOLD signal undershoot (right panel) to resolve more quickly, while the shallower slope (dashed line) leads to a much longer undershoot.

5. Because of the "stiffness" of the balloon during the onset of the flow change, the volume increases more slowly, and the change in deoxyhemoglobin is dominated by the change in blood oxygenation. The blood oxygenation reaches its new steady-state value before the volume has reached its new steady-state value, and during this gap, the total deoxyhemoglobin reaches its minimum value, corresponding to the initial overshoot in the BOLD signal. Note that in the context of the balloon model, such pronounced transient effects can occur in the presence of tight coupling of flow and oxygen metabolism.

DISCUSSION

In recent years, studies of the local changes in deoxyhemoglobin content associated with brain activation have found several transient effects that may be evidence for uncoupling of cerebral blood flow and oxygen metabolism. These features include initial dips and overshoots and a prolonged post-stimulus undershoot in the BOLD but not the flow signal. The earliest attempt to account for these transient features of the BOLD signal time course with a biophysical model that we are aware of is the work of Davis *et al.* (17). They modeled the capillary as a distributed compartment and the venous vasculature as a single well-mixed compartment. Model calculations showed an initial dip and for some combinations of parameters a post-stimulus undershoot, as well, although the time constant for the undershoot was shorter than what is usually observed. In the context of their model, O_2 is normally delivered in excess of what is needed for metabolism, so there is a resting non-zero O_2 concentration in the tissue. The initial dip is then due to an immediate increase of O_2 metabolism before flow increases by metabolizing this tissue pool of oxygen. Malonek and Grinvald (9) proposed a similar interpretation for the transient rise in deoxyhemoglobin. In the Davis model, the post-stimulus undershoot occurs when the increase in $CMRO_2$ is larger than the increase in delivery of O_2 by flow (the product of the unidirectional extraction and flow). Then, a larger fraction of the initially extracted O_2 is metabolized, so that the tissue concentration of O_2 drops. The post-stimulus undershoot then reflects continued extraction of O_2 to replenish the tissue oxygen pool after $CMRO_2$ has returned to normal. However, the explanation of the observed effects in terms of this model has two problems. First, the model does not appear to account for the long time constants observed for the undershoot. Second, blood volume changes are added on at the end, rather than being incorporated as an effect of flow changes. The Davis model also appears to be incompatible with the oxygen limitation model, which assumes that flow and O_2 metabolism are tightly coupled because all of the extracted O_2 is metabolized and there is no reserve pool of O_2 in tissue to draw on. This view is supported by the low tissue PO_2 values measured experimentally (30) and also by recent direct measurements of limited oxygen extraction (31).

To understand these various features of the BOLD time course, we have developed a biophysical model that is motivated by the Davis model but incorporates the oxygen limitation model (8) and blood volume changes as integral parts. The essential feature of the balloon model is that we assume that the capillary volume is fixed but the venous volume can change (like a balloon) with a pressure/volume response curve that may vary. In this initial development of the model, the shape of this curve is left as an unknown function, and Eq. [5] then follows as a consequence of mass balance. To connect the dynamic variables with measured BOLD signal curves, we used an expression for the MR signal based on recent theoretical calculations that should be appropriate for postcapillary vessels. Specifically including capillary effects, taking account of varying oxygen saturation along the capillary would improve the estimates of the constants k_1, k_2, and k_3. As these relationships become better understood through future experimental and theoretical work, these values should be revised, but the qualitative features of the model curves are not strongly sensitive to the precise values of the ks. The model curves shown in this article are meant to indicate the range of dynamic behavior that the balloon model can produce, and the

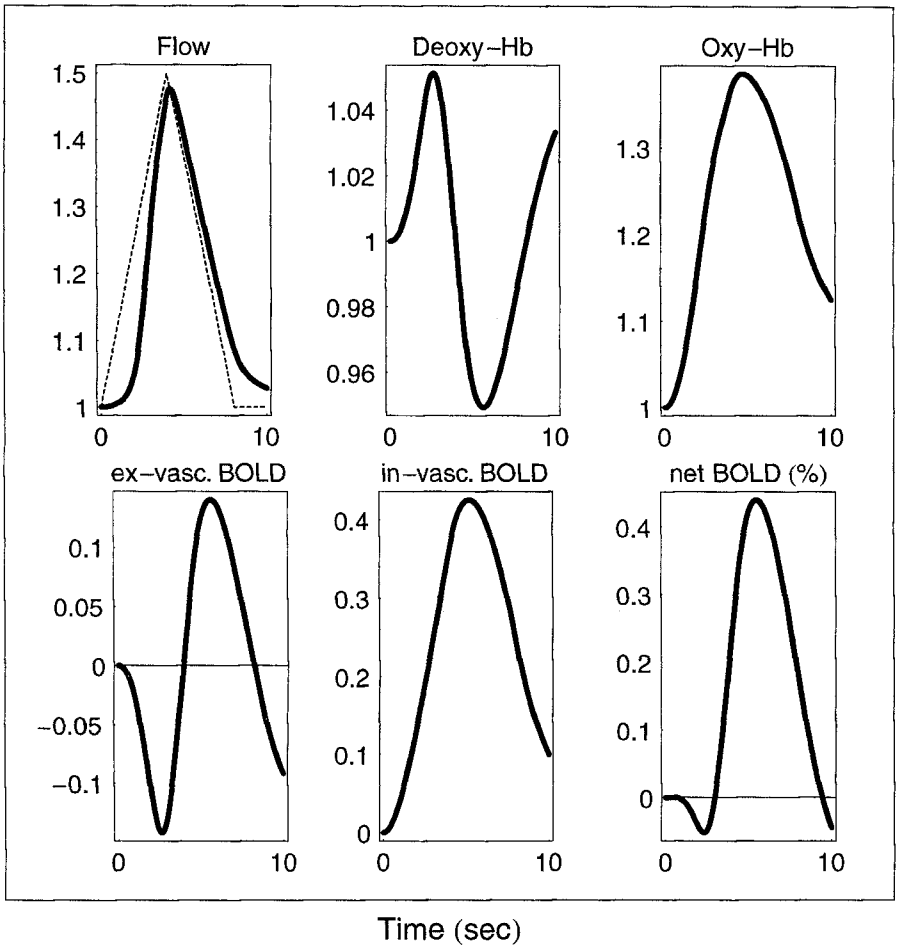

FIG. 5. The structure of the fast response in the deoxyhemoglobin, oxyhemoglobin, and BOLD signals. For this simulation, the in-flow increases linearly for 4 s then decreases linearly for 4 s. Outflow lags behind inflow by several seconds. In the bottom row, the intravascular, extravascular, and net BOLD signals are plotted separately. In this model, the initial increase in deoxyhemoglobin is due primarily to a blood volume increase, rather than a change in the concentration of deoxyhemoglobin in the blood. For this reason, only the extravascular component shows the initial dip. The conflict between the intravascular and extravascular signals reduces the dip in the net BOLD signal, which is a weighted sum of the two. These calculations are based on modeling of the intravascular signal for 1.5 T, where it contributes about two-thirds of the total signal change. At higher fields, the intravascular contribution is reduced relative to the extravascular signal, and so the initial dip should be more pronounced. Model parameters for these curves are the same as in Fig. 2.

precise shapes of the curves will vary with a different choice of parameters (e.g., a different shape of $f_{out}(v)$). Three basic patterns of BOLD signal transients are described by the model: (1) an initial dip and a corresponding increase in deoxyhemoglobin occur when inflow exceeds outflow and the blood volume rapidly increases; (2) an initial overshoot occurs when blood volume increases more slowly than inflow; and (3) a post-stimulus undershoot occurs when blood volume returns to baseline more slowly than inflow. In addition, the model predicts that the initial dip should be weak at 1.5 T because of conflicting signal changes in the extravascular and intravascular signals.

The balloon model is undoubtedly an overly simplistic biophysical model of the hemodynamic response to activation in the brain. This is in part due to the fact that the biomechanical properties of the brain vasculature are not well known. However, the model is a concrete mathematical representation of how blood volume changes in conjunction with blood oxygenation changes can create complex dynamics in the BOLD signal. One unexpected result was that the same model parameters that provide a good description of our BOLD data measured during finger tapping in humans (Figure 3), for a short duration stimulus, produce a pattern of deoxyhemoglobin change that is similar to previously published optical measurements in cat during visual stimulation (9) (Fig. 4). This similarity should be interpreted with caution, because the optical studies had a spatial resolution of 50 μm and large veins were excluded, whereas the fMRI data had a spatial resolution of several millimeters. Furthermore, the experimental data presented here only show the post-stimulus undershoot clearly; the temporal resolution and signal-to-noise ratio were not sufficient to reliably detect the fast response. Nevertheless, these comparisons with experimental data suggest that the model captures some of the essential features of the hemodynamic response.

The physical origin of these transient effects in the model is that the time constants for flow and blood volume changes are not the same, and recent experiments in animal models support this scenario (18, 19) for the post-stimulus undershoot. As illustrated here, a simple biomechanical model can produce blood volume changes that resolve very slowly, depending on the shape of the outflow curve as a function of volume. At this stage of development of the model, this curve is an arbitrary, but physically plausible, mathematical function. A natural extension of the model is a more complete biomechanical model of the viscoelastic properties of the vessels, so that the outflow curve could be expressed in terms of parameters related to these physical properties. Such a model would then allow estimates of these physical properties from measured flow and BOLD time courses.

These numerical calculations suggest that measurements of deoxyhemoglobin content are not reliable indicators of flow and oxygen metabolism changes without taking into account the volume changes as well. Pro-

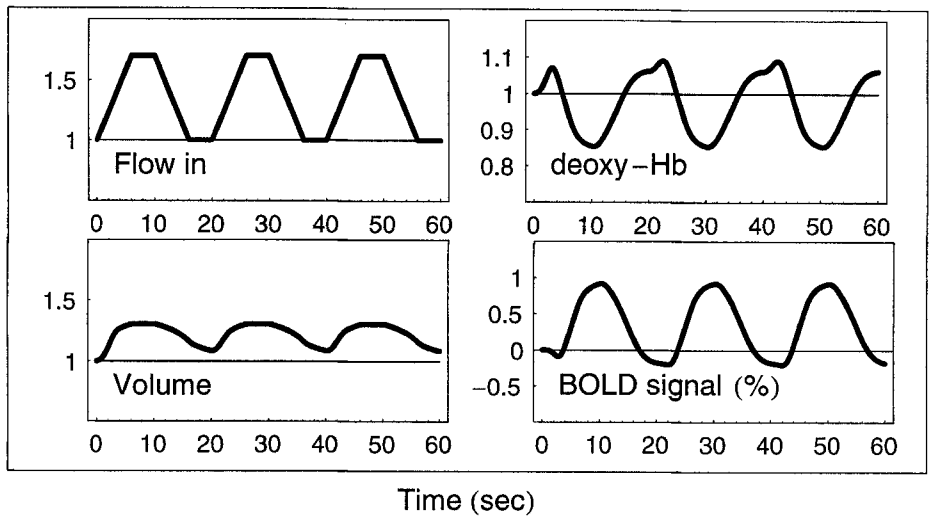

FIG. 6. Effects of multiple stimulus cycles. If the stimulus is repeated before the post-stimulus undershoot is resolved, the blood volume does not return to the resting value, and the apparent baseline in the BOLD signal time course is lowered. Note also that the shape of the deoxyhemoglobin response extracted from one cycle in the middle of a string of cycles is different from the initial cycle. For example, the initial cycle (t = 0–20 s) shows a large increase, followed by a decrease about three times larger, followed by an overshoot of the baseline. However, the second cycle (t = 20–40 s) appears to be an elevated baseline followed by a weaker increase, a strong decrease, and a return to the elevated baseline.

nounced transient features of the BOLD signal can result from differences in the separate time courses of flow and volume changes. Furthermore, these transient changes can occur even in the presence of tight coupling of flow and oxygen metabolism throughout the activation period.

APPENDIX

Derivation of the MR Signal Equation

We model the MR signal as a volume-weighted sum of the intrinsic extravascular signal S_e and the intravascular signal S_i, represented in Eq. [1]. A linear variation, valid for small changes, leads to an expression for the signal change, Eq. [2]. By factoring out V_0 and S_e from these equations, the fractional signal change can then be written as:

$$\frac{\Delta S}{S} = \frac{V_0}{1 - V_0 + \beta V_0} \cdot \left[\frac{1 - V_0}{V_0} \frac{\Delta S_e}{S_e} + \frac{\Delta S_i}{S_e} + (1 - v)(1 - \beta) \right] \quad [A1]$$

where $\beta = S_i/S_e$ is the intrinsic signal ratio at rest, and v is the blood volume normalized to the volume at rest. Ogawa *et al.* (23) found by numerical simulations that, for larger vessels, the gradient echo transverse relaxation rate is approximated by $R_2^* = 4.3$ fV, where $f = \Delta X \omega_0 E$ is the susceptibility difference between the intra- and ex-

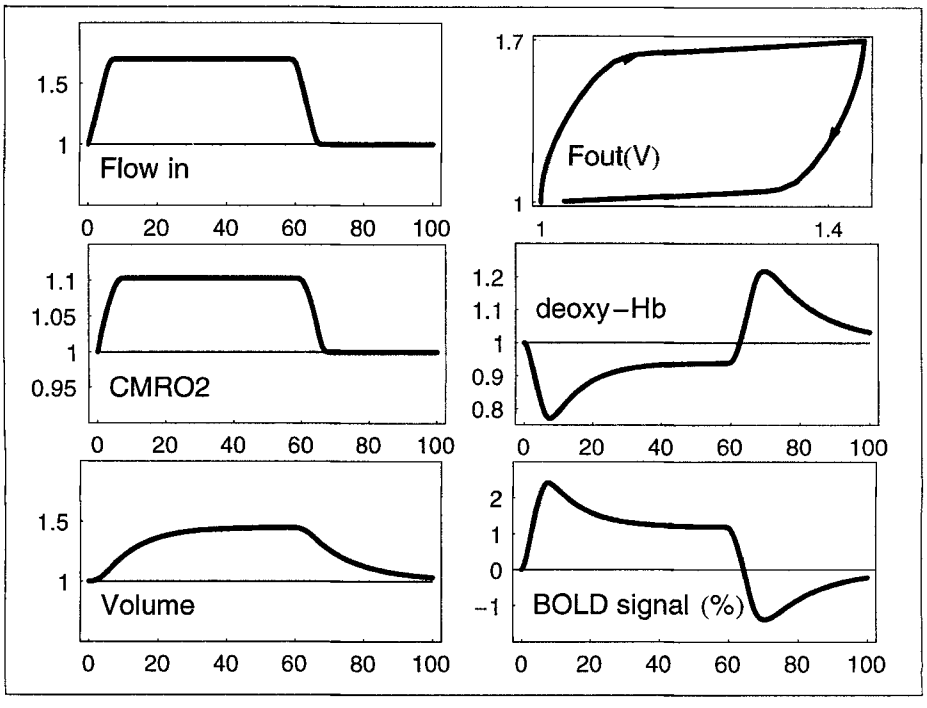

FIG. 7. The effect of hysteresis in the form of $f_{out}(v)$. In this calculation, the outflow as a function of v shown in the upper right was different on expansion and contraction (as indicated by the arrows), roughly approximating a balloon that is initially stiff, but after prolonged expansion is more compliant. In addition, the volume change in this example is larger to illustrate the type of effects that can arise. The BOLD signal shows a strong initial overshoot, followed by a gradual return to a reduced plateau, and ending with a strong post-stimulus undershoot, similar in pattern to experimental data reported by Frahm and coworkers (15, 16). Inflow and $CMRO_2$ are coupled throughout and show a simple trapezoidal pattern. Additional model parameters in these calculations were $E_0 = 0.4$, $V_0 = 0.02$, and $\alpha = 0.7$.

travascular spaces expressed as a frequency, $V = vV_0$ is the blood volume fraction, and E is the O_2 extraction fraction. To put this into our notation with total deoxyhemoglobin q and volume v normalized to their values at rest, it is helpful to note that $E/E_0 = q/v$. Then the transverse relaxation rate is $R_2^* = 4.3\ \Delta X\omega_0 E_0 V_0 q$. The change in the extravascular signal is then:

$$\frac{\Delta S_e}{S_e} = -\Delta R_2^* TE = aV_0(1 - q) \qquad [A2]$$

with $a = 4.3\ \Delta X\omega_0 E_0 TE$. For 1.5 T and $\Delta X = 1$ ppm, $\Delta X\omega_0 = 40.3\ s^{-1}$, and with resting extraction $E_0 = 0.4$ and TE = 40 ms, the scaling factor is approximately $a = 2.8$.

For the intravascular signal change and the resting signal ratio β, we used the results of Boxerman *et al.* (25). In their numerical simulations for 1.5 T, TE = 40 ms, and vessel radius = 25 μm, they calculated the attenuation factor A of the intravascular component as a function of the fractional oxygen saturation of hemoglobin. Based on their results (their Fig. 2), the attenuation factor can be approximated as $A = 0.4 + 2(0.4 - E)$ in the range $0.15 < E < 0.55$. Assuming that the full intra- and extravascular signals are the same in the absence of deoxyhemoglobin, the resting signal ratio is $\beta = A$, so for $E_0 = 0.4$, $\beta = 0.4$. The intravascular signal change is then:

$$\frac{\Delta S_i}{S_e} = -\Delta A = -2\Delta E = 2\left(1 - \frac{q}{v}\right) \qquad [A3]$$

Combining Eqs. [A1] to [A3] yields Eq. [3], and because the factor multiplying Eq. [A1] is approximately V_0 for small resting blood volume fractions, the coefficients are $k_1 \cong a \cong 2.8$, $k_2 \cong 2$, and $k_3 \cong 0.6$.

REFERENCES

1. P. T. Fox, M. E. Raichle, Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proc. Natl. Acad. Sci. USA* **83,** 1140–1144 (1986).
2. P. T. Fox, M. E. Raichle, M. A. Mintun, C. Dence, Nonoxidative glucose consumption during focal physiologic neural activity. *Science* **241,** 462–464 (1988).
3. K. K. Kwong, J. W. Belliveau, D. A. Chesler, I. E. Goldberg, R. M. Weisskoff, B. P. Poncelet, D. N. Kennedy, B. E. Hoppel, M. S. Cohen, R. Turner, H.-M. Cheng, T. J. Brady, B. R. Rosen, Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Natl. Acad. Sci. USA* **89,** 5675–5679 (1992).
4. S. Ogawa, D. W. Tank, R. Menon, J. M. Ellermann, S.-G. Kim, H. Merkle, K. Ugurbil, Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA* **89,** 5951–5955 (1992).
5. P. A. Bandettini, E. C. Wong, R. S. Hinks, R. S. Tikofsky, J. S. Hyde, Time course EPI of human brain function during task activation. *Magn. Reson. Med.* **25,** 390–397 (1992).
6. K. K. Kwong, Functional magnetic resonance imaging with echo planar imaging. *Magn. Reson. Q.* **11,** 1–20 (1995).
7. J. W. Prichard, B. R. Rosen, Functional study of the brain by NMR. *J. Cereb. Blood Flow Metab.* **14,** 365–372 (1994).
8. R. B. Buxton, L. R. Frank, A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *J. Cereb. Blood Flow Metab.* **17,** 64–72 (1997).
9. D. Malonek, A. Grinvald, Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: implications for functional brain mapping. *Science* **272,** 551–554 (1996).
10. P. A. Bandettini, E. C. Wong, J. R. Binder, S. M. Rao, A. Jesmanowicz, E. A. Aaron, T. F. Lowry, H. V. Forster, R. S. Hinks, J. S. Hyde, *in* "Diffusion and Perfusion: Magnetic Resonance Imaging" (D. LeBihan, Ed.), pp. 335–349, Raven Press, New York, 1995.
11. T. Ernst, J. Hennig, *Observation of a fast response in functional MR. Magn. Reson. Med.* 146–149 (1994).
12. R. S. Menon, S. Ogawa, J. P. Strupp, P. Anderson, K. Ugurbil, BOLD based functional MRI at 4 tesla includes a capillary bed contribution: echo-planar imaging correlates with previous optical imaging using intrinsic signals. *Magn. Reson. Med.* **33,** 453–459 (1995).
13. X. Hu, T. H. Le, K. Ugurbil, Evaluation of the early response in fMRI in individual subjects using short stimulus duration. *Magn. Reson. Med.* **37,** 877–884 (1997).
14. T. L. Davis, R. M. Weisskoff, K. K. Kwong, R. Savoy, B. R. Rosen, Susceptibility contrast undershoot is not matched by inflow contrast undershoot, *in* "Proc., SMR, 2nd Annual Meeting, San Francisco, 1994," p. 435.
15. J. Frahm, G. Krüger, K.-D. Merboldt, A. Kleinschmidt, Dynamic uncoupling and recoupling of perfusion and oxidative metabolism during focal brain activation in man. *Magn. Reson. Med.* **35,** 143–148 (1996).
16. G. Kruger, A. Kleinschmidt, J. Frahm, *Dynamic MRI sensitized to* cerebral blood oxygenation and flow during sustained activation of human visual cortex. *Magn. Reson. Med.* **35,** 797–800 (1996).
17. T. L. Davis, R. M. Weisskoff, K. K. Kwong, J. L. Boxerman, B. R. Rosen, Temporal aspects of fMRI task activation: dynamic modeling of oxygen delivery, *in* "Proc., SMR, 2nd Annual Meeting, San Francisco, 1994," p. 69.
18. J. B. Mandeville, J. Marota, J. R. Keltner, B. Kosovsky, J. Burke, S. Hyman, L. LaPointe, T. Reese, K. Kwong, B. Rosen, R. Weissleder, R. Weisskoff, CBV functional imaging in rat brain using iron oxide agent at steady state concentration, *in* "Proc., ISMRM, 4th Annual Meeting, New York, 1996," p. 292.
19. J. B. Mandeville, J. J. A. Marota, B. E. Kosofsky, J. R. Keltner, R. Weissleder, B. R. Rosen, R. M. Weisskoff, Dynamic functional imaging of relative cerebral blood volume during rat forepaw stimulation. *Magn. Reson. Med.* **39,** 615–624 (1998).
20. R. B. Buxton, E. C. Wong, L. R. Frank, A biomechanical interpretation *of the BOLD signal time course: the Balloon Model, in* "Proc., ISMRM, 5th Annual Meeting, Vancouver, 1997," p. 743.
21. R. M. Weisskoff, C. S. Zuo, J. L. Boxerman, B. R. Rosen, Microscopic susceptibility variation and transverse relaxation: theory and experiment. *Magn. Reson. Med.* **31,** 601–610 (1994).
22. C. R. Fisel, J. L. Ackerman, R. B. Buxton, L. Garrido, J. W. Belliveau, B. R. Rosen, T. J. Brady, MR contrast due to microscopically heterogeneous magnetic susceptibility: numerical simulations and applications to cerebral physiology. *Magn. Reson. Med.* **17,** 336–347 (1991).
23. S. Ogawa, R. S. Menon, D. W. Tank, S.-G. Kim, H. Merkle, J. M. Ellerman, K. Ugurbil, 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 (1993).
24. D. A. Yablonsky, E. M. Haacke, Theory of NMR signal behavior in magnetically inhomogenous tissues: the static dephasing regime. *Magn. Reson. Med.* **32,** 749–763 (1994).
25. J. L. Boxerman, P. A. Bandettini, K. K. Kwong, J. R. Baker, T. L. Davis, B. R. Rosen, R. M. Weisskoff, The intravascular contribution to fMRI signal change: Monte Carlo modeling and diffusion-weighted studies in vivo. *Magn. Reson. Med.* **34,** 4–10 (1995).
26. R. L. Grubb, M. E. Raichle, J. O. Eichling, M. M. Ter-Pogossian, The effects of changes in PCO_2 on cerebral blood volume, blood flow, and vascular mean transit time. *Stroke* **5,** 630–639 (1974).
27. E. C. Wong, L. R. Frank, R. B. Buxton, QUIPSS II: a method for improved quantitation of perfusion using pulsed arterial spin labeling, *in* "Proc., ISMRM, 5th Annual Meeting, Vancouver, 1997," p. 1761.
28. E. C. Wong, R. B. Buxton, L. R. Frank, Implementation of quantitative perfusion imaging techniques for functional brain mapping using pulsed arterial spin labeling. *NMR Biomed.* **10,** 237–249 (1997).
29. E. C. Wong, P. A. Bandettini, J. S. Hyde, Echo-planar imaging of the human brain using a three axis local gradient coil, *in* "Proc., SMRM, 11th Annual Meeting, Berlin, 1992," p. 105.
30. D. W. Lubbers, H. Baumgartl, W. Zimelka, Heterogeneity and stability of local PO2 distribution within the brain tissue. *Adv. Exp. Med. Biol.* **345,** 567–574 (1994).
31. I. G. Kassissia, C. A. Goresky, C. P. Rose, A. J. Schwab, A. Simard, P. M. Huet, G. G. Bach, Tracer oxygen distribution is barrier-limited in the cerebral microcirculation. *Circ. Res.* **77,** 1201–1211 (1995).