Chapter 1

Neural Signal Estimation through Time-Resolved Functional Imaging

CW Tyler and LT Likova.

Abstract

The most effective method for localizing the effects of neural activation throughout the human brain is Functional Magnetic Resonance Imaging (fMRI). However, there is currently no method able to derive from the fMRI signal the particular neural signals underlying the measured BOLD waveforms. The goal of the present analysis is therefore to achieve time-resolved estimation of the neural signals determining the temporal waveforms in event-related fMRI. This analysis provides a new level of specificity in the basis of advanced approaches of using fMRI to study brain function as deep as at the level of the neural processing. In particular, we can resolve details of neural response dynamics down to the millisecond level for the analysis of visual processing and its temporal deficits.

For this purpose, we develop a nonlinear dynamic forward model of the metabolic demand arising from the transmitter release in the minimal set of component neural populations that account for each fMRI waveform. This model allows derivation of the neural signals underlying the wide variety fMRI waveforms evoked from even the most elementary brief stimuli. Estimation of signal delays between cortical regions is enhanced by an autonormalization procedure to provide maximum temporal precision in mapping the causal interactions among functional subunits across the cortex by means of the pattern of simultaneous covariation in the fMRI responses. This technique is also robust to feedback processes, allowing the mapping of recursive feedback signal between cortical regions at different levels in the response hierarchy.

Girard IC, Andre JS (Eds), Brain Mapping Research Progress, NSP: New York, 2009.

Overview

The basic capability of functional Magnetic Resonance Imaging (fMRI) is mapping the spatial organization of the cortical and subcortical responses throughout the brain to a variety of stimulus and behavioral paradigms. It is often considered to have high spatial resolution (0. 25 – 1 million voxels), but low temporal resolution (0.5 – 5 *s*) relative to other methods for mapping human brain function (such as high-density EEG analysis). Of course, even the spatial resolution is low relative to single-unit neurophysiological recoding, but this is largely attributable to the signal/noise ratio. Inplane spatial resolution as high as 150 μ has been reported at 9.4 Tesla in cat cortex (Harel et al., 2006) and less than 1 mm^3 voxels in human (Barth and Norris, 2007). As improved technologies come on line, these spatial limitations are expected to be continually decreased.

The temporal limit is not so tractable. The inherent time constants of the BOLD signal are about 5 *s*, several orders of magnitude slower than the neural signal that is driving it. This slow time course incurs a severe limitation on the use of fMRI for application to deficits in temporal processing and diseases such as multiple sclerosis, where the conductions delays are of the order of tens of *ms* (Regan, Milner and Heron, 1976; Neima & Regan, 1984; Stefano et al., 1991; Werring et al., 2000; Russ et al., 2002; Levin et al., 2006). For a reasonable recording time and voxel size (*e. g.*, 3 x 3 x 3 mm) the signal/noise ratio for the most effective stimuli rarely exceeds 10:1. At such signal/noise ratios, deconvolution of the BOLD response can improve its temporal resolution by about a factor of 5 (Glover, 1999; Logothetis, 2003). However, that for the underlying neural signal can only be improved to about 1 sec, which is far short of what is required to measure typical neural delays. Single-parameter estimation, such as response delay, can improve the temporal resolution to 100 *ms* or better for narrowly targeted brain regions (Menon, Luknowsky and Gati, 1998; Henson et al., 2002), but this requires the assumption that the BOLD signal has a unitary waveform, which is often not the case. Even minor deviations from a stable waveform violate the assumptions of such single-parameter analysis and invalidate the delay measure.

Indeed, most commonly used fMRI analysis techniques, such as the SPM analysis package, make the assumptions of a unitary BOLD waveform kernel for an instantaneous stimulus, which, together with the assumption of superposition linearity, allows the prediction of the BOLD response waveform for any stimulus type or duration in any brain area. In fact, however, major deviations from a standard BOLD waveform may be found, even in the same cortical regions, for variations in stimulus conditions. D'Avossa, Shulman and Corbetta (2003) reported strong differences in waveform for attention modulation of response to the motion or color of a cue/stimulus pairing with about 5 *s* duration. Likova and Tyler (2007) showed striking differences within each of many cortical regions for manipulation of the perceptual organization in result of instantaneous changes in random-noise field. A similar variety of BOLD waveforms for a set of visual stimuli within given cortical regions are reported in a later section of the present paper. Such local waveform differences must derive from

differences in the *neural* signals driving the BOLD activation, since the metabolism and hemodynamics should be invariant within a given cortical region. These results imply that the fMRI signal contains much more information than commonly appreciated. It is therefore important to develop a coherent analysis of the neural-to-BOLD processing sequence in order to derive a clear understanding of the neural signal variation revealed by the BOLD variations.

There is at present no method of transcending the BOLD temporal limitations in order to estimate the neural signals underlying the measured fMRI waveforms. The present analysis is therefore designed to provide *time-resolved estimation* of the *neural signals* underlying the particular characteristics of the temporal BOLD waveforms for a particular stimulus processed by a particular cortical region. We propose a *Nonlinear Forward Modeling* approach to provide a compact account of the measured waveform with the minimal number of neural predictors. This approach utilizes prior knowledge of the expected temporal properties of neural signals and of their consequent metabolic demand. This approach is reminiscent of the spectral analysis of the composition of stars, in which each element has a characteristic pattern of emission lines and the net spectrum is the sum of an unknown mixture of these predetermined patterns. The result is a specification of prevalence of each element in the star. In the case of the neural signals, the result is that the amplitude and delay of each of the neural components that, when summed, account for the measured waveform for a particular stimulation condition and cortical region. The predictors are non-linear because there is a nonlinear relationship between the neural response and the metabolic demand that it generates, but we can assume that the component metabolic demands sum linearly together.

The next step in neuroscience, beyond spatial and temporal mapping, is the establishment of the causal relations among the network of activated cortical and subcortical locations. We introduce an autonormalization procedure to provide maximum temporal precision in mapping the functional subunits across the cortex by means of pattern of simultaneous covariation. The causal mapping of such delays can then be tracked through responsive cortical and subcortical regions to determine the causal networks with unprecedented accuracy. The technique is also robust to feedback processes, allowing the mapping of recursive feedback signal between regions lower and higher in the response hierarchy (when the delay is sufficiently long to disambiguate the feedback signals from the feedforward projections).

The Chain of Processing

In general terms, the stimulus S(t) impinging on the subject generates a sequence of neural responses starting with the transduction into a neural signal within the sensory receptors, which then propagates to the brain and activates various populations of neurons within the voxels being analyzed by the fMRI technique. For instance, the signals arriving from the retina generate synaptic activation of the populations of cortical cells, which generates an energetic demand for the restoration of the neurotransmitter molecules carrying the activation signals. The chain of cortical processing, illustrated in the block diagram of Fig. 1, progresses from the neural events

through the metabolic demand mediated by the neighboring astrocyte glial cells to the processes of oxygen delivery by the adjacent capillaries that is detected by the imaging methodology. At each stage, the primary signal (denoted by capital symbols) is considered to be generated by convolution of the signal from the previous stage with an impulse response kernel (denoted by lower-case symbols) specifying the temporal response of the current stage. At particular levels characteristic nonlinearities are introduced before proceeding to the next stage. These convolutions are specified in Table I but are omitted from the following text for simplicity.

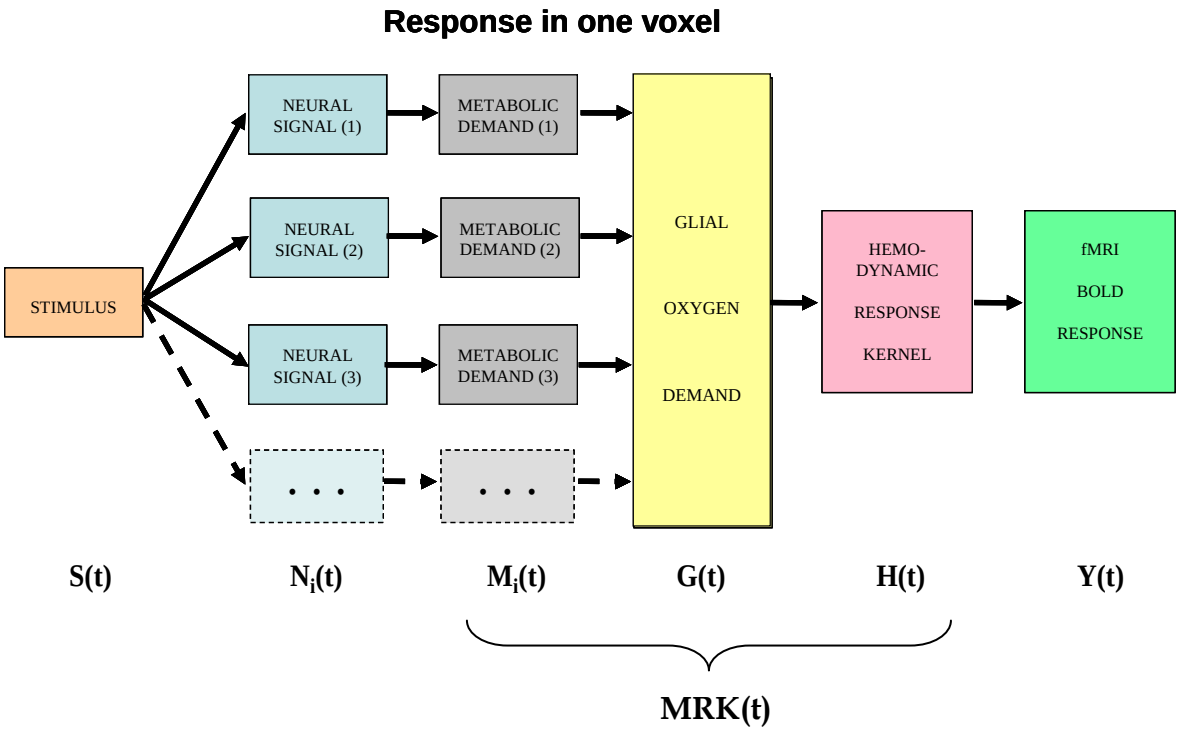

Fig. 1. Block diagram of the main processing stages that lead up to the BOLD signal. The subscript i indicates that the stage incorporates multiple components within the voxel.

We will treat the *neural responses* within a given voxel as generated by sets of homogeneous populations with similar signal waveforms $N_i(t)$ within each *i*th population. Each neural response then generates a *local metabolic demand* $M_i(t)$ that may have a nonlinear relationship to the neural signal waveform. The *integrated metabolic demands* $G(t)$ are met primarily by the astrocytes, which integrate the required energy consumption over time and space and make a *complementary oxygen demand* $G(t)$ on the adjacent vasculature. The *hemodynamic processes* $H(t)$ provide the requisite oxygen exchange to replenish the energy depletion in the astrocytes. The fMRI BOLD analysis provides an estimate $Y(t)$ of the ratio of oxygenated to deoxygenated hemoglobin in the blood complement of a given voxel. The *post-neural* processing stages are often modeled as a *linear hemodynamic response kernel* convolved with the presumed neural signal. However, this approach overlooks the key role of the *pre-hemodynamic* processes of the *glial* intermediaries. To incorporate the contributions of these processes, we will term it

a *'metabolic response kernel'* (MRK) incorporating *both* the glial and hemodynamic components of the metabolic recovery processes. The MRK will be convolved with a non-linear transform of the presumed neural signal to provide an estimate of the neural metabolic demand that is being met by the combined glial and hemodynamic metabolic response. The fMRI analysis also has a finite dynamic response time, but it will be treated as negligible and linear with respect to the MRK of the glial/hemodynamic response.

Table I. Mathematical model of the operations involved in the generation of the BOLD signal from the input stimulus, S(t).

Output	**Generation logic**	**Features**
Neural signal	$N_i(t) = f[S(t) \otimes n_i(t)]$	Nonlinear transducer
Neural metabolic demand	$M_i(t) = f[N_i(t) \otimes m(t)]$	Nonlinear transducer
Glial oxygen demand	$G(t) = \sum M_i(t) \otimes g(t)$	Multiple linear integrator
Hemodynamic response	$H(t) = G(t) \otimes h(t)$	Linear (slow)
Paramagnetic response	$BOLD(t) = H(t) \otimes p(t)$	Linear (fast)
	$\approx \sum M_i(t) \otimes MRK(t)$	Combines four linear stages

The terms in Fig. 1 are related by a series of mathematical operations specified in Table I. The three operators are: (i) linear convolution (⊗), a nonlinear, zero-memory transducer ($f[\ ]$) and a linear integrator ($\sum$). In this model, which is considerably elaborated from the conventional analyses, we make the simplification that each stage consists of the linear convolution of the output signal from the previous stage with a temporal response kernel (which is designated by lower case initial for the name of that stage). Where necessary to model the physiological processes, the linear convolution kernel is elaborated with one of the nonlinear operators.

To pursue the analysis of neural responses underlying BOLD signals, we consider the responses to rapid stimulus events in an event-related paradigm. The measured event-related BOLD response *Y(t)* for this paradigm will be termed the *Peri-Event Response Function* (PERF), to distinguish it from *h(t)*, the underlying *hemodynamic response function* (HRF), or *kernel*. (The HRF is the kernel specific to the dynamics of blood oxygenation and is based on the assumption that its input is an invariant Dirac pulse.) The PERF is thus defined as the averaged BOLD activity synchronized with any repeated stimulus event. It incorporates the full sequence of linear or nonlinear neural, metabolic/hemodynamic response properties that follow the stimulus event of

whatever form. Unlike the HRF, the PERF concept does not make any assumptions about the stimulus or the mechanism, location or linearity of the measured responses.

Time-Resolved fMRI

Although the time resolution of FMRI is low relative to the full time resolution of single unit activity, it must be recognized that many single-unit studies are noise-limited and use response integration up to 500 *ms* or so, and that many interesting perceptual processes have durations and delays this long and longer. Thus, resolving fMRI delays in the range below 1 sec can deliver important information, as demonstrated by Menon et al. (1998). These authors were able to resolve reaction-time delays of the order of 100 *ms*, and Henson et al. (2002) used the basis function of the HRF derivative to determine the sign of the advance or delay of the BOLD response (although the latter approach assumes a unitary HRF in all response regions, which may confound delays with other waveforms differences (Bellgowan, Saad and Bandettini, 2003). Similarly, Sun et al. (2005), working with a frequency domain technique, used the slope of the phase of the fMRI spectrum as an index of relative neural delay between brain areas. This method allowed the mapping of causal influences from a chosen brain areas to the rest of the brain with a typical accuracy of about 500 *ms*.

The present analysis is focused on the *time* domain, with the goal of extracting the maximum possible information about the neural signals underlying particular BOLD activation profiles. It is important to realize that there are many situations in which the neural response properties are reflected in the BOLD signals, even if the time-resolution is insufficient to reproduce the exact neural signal. Examples of such 'transparency' of response properties are particularly clear in the case that the full metabolic response kernel is monophasic.

This seems to be a fair approximation in the case of fMRI because the canonical response kernel (commonly known as the HRF) provided in the SPM-5 software package, although biphasic, has a negative lobe of less than 10% of the amplitude of the positive lobe. It is therefore only a minor approximation of this standardized kernel to assume that it has no negative lobe, which is the assumption made for the following analysis. (As will be explained below, the residual biphasic component can be readily attributed to neural rebound signals.) Armed with this assumption, we show how several properties of the neural signal are reflected in the recorded PERF waveform. (For the purpose of this demonstration, we are assuming a linear relationship between the neural response and the PERF in order to make its properties clear before introducing the nonlinear aspects discussed in the previous section.)

Polarity. The PERF polarity will be an accurate reflection of the polarity of the neural response for any monophasic neural response.

Latency. Any delay in the neural response will also be reflected in the consequent PERF. Of course, the metabolic processing sequence may introduce additional delays, but neural delays such as response reaction times or perceptual ambiguity delays should be accurately reflected in the PERF once the inherent delays of the MRK are taken into account.

Transience. It is an established property of the BOLD response that it is largely sustained for an appropriately sustained neural response (Birn, Saad and Bandettini, 2001; Boynton et al., 1996; Glover, 1999; Logothetis, 2002, 2003; Shmuel et al., 2006). Hence, transience of the PERF for a sustained stimulus implies a transience of the underlying neural response. For example, the onset of a sustained light is known to generate transient neural responses in most cortical neurons responding to it. Such stimulation will generate a transient PERF even though the stimulation and photoreceptor response are sustained.

Number of phases. For a monophasic MRK (HRF), the PERF will have the same number of phases as the input stimulus, if the neural input is balanced for positive and negative lobes. Thus, the fact that the measured PERF is typically biphasic does not imply that the MRK is necessarily biphasic. The negative lobe may derive from a biphasic neural response to a stimulus rather than to blood dynamics.

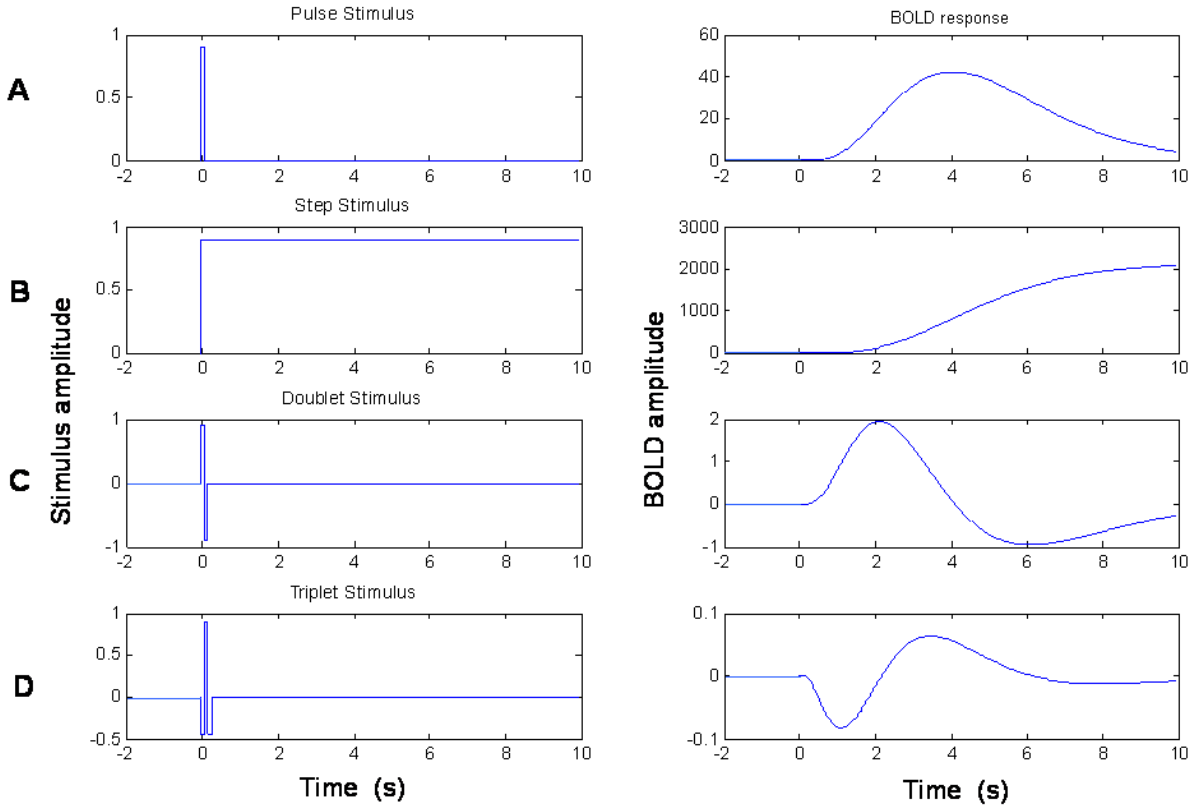

Fig. 2. Left panels: (***A***) neural impulse response, (***B***) step response, (***C***) balanced doublet and (***D***) balanced triplet response. Right panels present convolution of each of these responses with the MRK shown in *A* (right). Note that differences in neural response characteristics (left panels) at the time scale of 100 *ms* generate profound changes in the simulated PERF waveforms (right panels) on a much longer timescale, which in turn are diagnostic of the differences in the neural signals.

These properties of the PERF are illustrated in the simulation of Fig. 2, for which the MRK (HRF) is assumed to be the gamma function shown in Fig. 2***A***, right. (We use the term "gamma function" as shorthand for the gamma probability density function

that is the theoretical limit of the Gamma Distribution, which is the sum of k exponentially-distributed random variables). Gamma probability density functions have the form $a.\ t^{k-1}.\ e^{-t/\tau}$ (where t is the time dimension, a is a scaling parameter and k and τ are shape constants). Each left-hand panel shows a different neural waveform with a time-course typical of those seen in neurophysiological recordings. Convolution with the MRK of Fig. 2*A* (right) generates the predicted BOLD waveforms of the right-hand column of Fig. 2. The important point made by this simulation is that most typical features of measured PERF waveforms could arise from convolution of a simple gamma-function MRK with a variety of neural signals modulated within less than 100 *ms*.

Conversely, under the assumption of a uniform MRK, the neural signal properties can be inferred from the form of the PERF. Thus, if the PERF rises to a sustained plateau (Fig 2***B***, right), we may infer that the neural signal was sustained (Fig 2**B**, left). If the PERF is biphasic (Fig 2***C***, right), we may infer that the neural signal was biphasic (Fig 2***C***, left). If the PERF is triphasic (Fig 2***D***, right), we may infer that the neural signal was correspondingly triphasic (Fig 2***D***, left). Fig. 2 illustrates that, under appropriate assumptions the form of the BOLD response can act as a kind of "temporal microscope" for the form of the underlying neural response. Although it cannot resolve the precise temporal parameters (*e. g.*, whether the response lobes have durations of 1, 10 or 100 *ms*), many of the aspects of the waveform structure are reflected in the BOLD waveform. The interpretation is particularly straightforward if the BOLD metabolic/hemodynamic kernel is monophasic, and we are emphasizing that previous suggestions that it is biphasic are contaminated by the likelihood that the underlying neural signal is itself biphasic. Under the logic of the analysis in Fig. 2, this property would generate the biphasic form of empirical hemodynamic kernel that is commonly observed, but it seems that no previous analysis of BOLD properties has taken this possibility into account. Based on this analysis, we propose that (i) it is more plausible to assume that the true metabolic kernel has a monophasic form, and that (ii) the biphasic property that is often empirically observed (but is quite variable over brains and cortical regions; Bellgowan, Saad & Bandettini, 2003; Handwerker et al., 2003), is a function of the neural processing that precedes the metabolic demand, as illustrated in

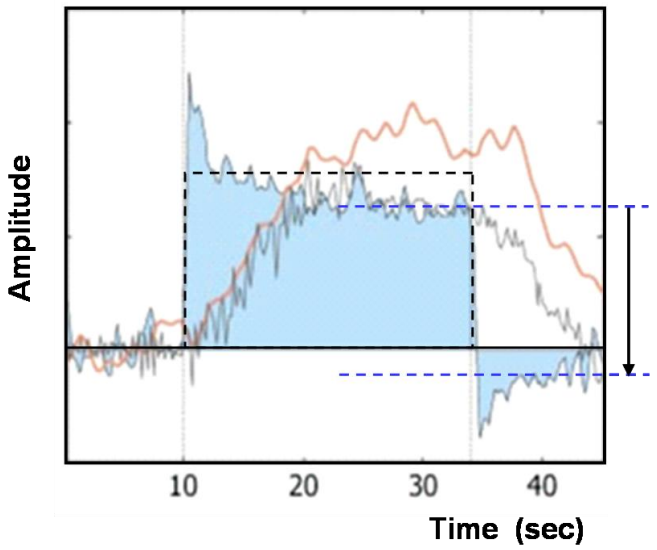

Fig. 3. Time course of the local field potentials (blue trace), BOLD (red trace) and predicted BOLD (black trace) to a continuous dynamic stimulus (black dotted rectangle) of 24 s duration. From Logothetis (2003). Blue dashed lines show the inhibitory rebound level predicted from the average level of the second half of the BOLD response.

Fig. 2.

The Neural Signal

Logothetis and colleagues (Logothetis, 2002, 2003; Logothetis & Wandell, 2004; Shmuel et al., 2006) have made the case that the most likely basis of the metabolic demand driving the BOLD signal is the energetic load deriving from the total conductance changes in the postsynaptic membrane generated by a range of processes consequent on transmitter release at the synaptic inputs to each neuron. The summed metabolic demand in the nexus of active cortical neurons adjacent to a capillary forms the drive for the metabolic response in that region of cortex. These authors argue that the transmitter release is tightly coupled to the activation of the post-synaptic receptors on the recipient cell membrane and consequently to the energetic demands of the membrane receptor activation and to a lesser extent to the subsequent recycling of the transmitter molecules. The majority of these energetic demands are met by glycolysis of glutamate to glutamine in the neighboring astrocytes (Shank and Aprison, 1979; Wang and Floor, 1994). The glutamine is then taken up by the neurons for reconversion to glutamate for use as a transmitter, releasing energy within the neuron in the process. Since cells are predominantly linear summators of the excitatory and inhibitory transmitter release across their synaptic population, it seems to follow that the energetic demand driving the BOLD signal is most closely coupled to the net transmitter signal impinging on the cells, and hence to intracellular potential in the cells.

A comprehensive model of the BOLD therefore requires an accurate model of the intracellular potential dynamics coupled to sensory stimulation. As is highlighted by the data of Logothetis (2003), there are adaptive effects in the neural response that become evident for long-duration activation by continuous dynamic stimuli (Fig. 3). The LFP recordings (blue curve) from monkey V1 reveal that the adaptation has a complex time course that can be approximated by two exponentials with time constants of about 1 sec and 60 sec, respectively. These are remarkably prolonged neural processes that are on the time scale of BOLD activation, and will therefore affect the form of the recorded BOLD signal from the same general region of cortex (Fig. 3, black trace). The importance of this adaptation effect is emphasized by the fact that the recorded LFP signal does not fully match the BOLD activation predicted (red curve, Fig. 3) on the basis of General Linear Model (GLM) of convolution of a hemodynamic response kernel with the stimulus time course, and therefore a more comprehensive model is required, going beyond the standard GLM model. (This analysis applies when the stimulus is continuously refreshed to support the assumption that the neural signal would be well approximated by the stimulus waveform).

Notice that the negative LFP signal following the stimulus *offset* in Fig. 3 has a similar (but inverted) time course relative to that following the stimulus *onset*. This inversion implies that the adaptation effect is a subtractive *inhibition* (rebound prediction shown by the blue dashed lines) rather than solely a multiplicative form of fatigue (which would have no negative rebound). If it were a multiplicative gain

control, the amplitude of signal change at offset would be substantially less than that at onset, whereas the two amplitudes are similar within about 10%.

The mechanism of this subtractive inhibition appears to be the tonic intracellular hyperpolarization induced by pattern adaptation, as described by Carandini and Ferster (1997, 2000). However, the stimulus driving this response was a *dynamic* contrast modulation, and hence this gives us a solid basis to treat the sustained LFP signal as deriving from a full-wave rectified transform of the intracellular potential. This leads to a model of the LFP from any nearby neuron as the integrated response of instantaneous signal of the rectified intracellular potential. As developed by Aubert & Costelet (2002), the initial response of the sodium pump to the influx of transmitter molecules due to stimulus activation has a nonlinaear relaxation term that may be modeled with an dual exponential decay function with a short decay to a stable response level followed by a slow decay of about 10% over the 30 sec of the full stimulus epoch. These components represent the effect of two separate forms of shunting inhibition contributing to the recorded LFP. Note that it is difficult to obtain the integrated response, i.e., the combination of the two component slopes of Fig. 4**c** with any kind of serial model, because this would imply a convolution of the two exponentials that would necessarily result in a function dominated by the faster process rather than allowing both processes full expression.

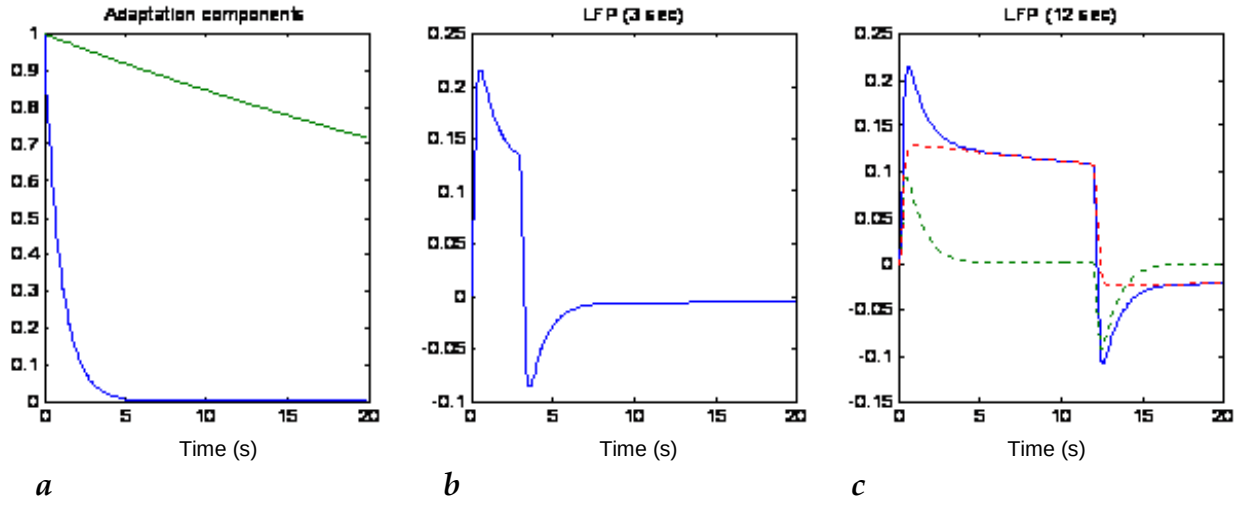

Fig. 4 **a**) Two exponential decay processes used to account for the adaptation effects in Fig. 3. ***b***) The sum of the two components in 4*a* each convolved with a 3 sec stimulus time course. c) *Dashed curves*: the two components in 4*a* each convolved with a 12 sec stimulus time course; *full curve*: the sum of the two convolved responses.

Temporal Properties of Human Vision

To place the temporal analysis of human visual processing on a firm footing, it is critical to have a secure understanding of the linearity of the signals passing from the retina to the cortex that form the basis of visual perception. This knowledge may be achieved by psychophysical measurements of the temporal integration function (Bloch, 1885;

Barlow, 1958). An example of such measurements for three base intensities of red (660 nm) light at the eccentricity of 35° is depicted in Fig. 5, which shows three properties of relevance. One is that, below 10 *ms* duration, detection thresholds have an accurately reciprocal relation to pulse duration down to durations as low as 10 μs (at the higher mean intensities). This behavior implies that temporal summation is accurately linear over this extreme range. The number of quantal absorptions (q) per cone implied by the reciprocal relationship is indicated next to each curve. If, on the other hand, the cone or subsequent neural responses had an amplitude saturation, for example at a level of 1 million quanta (6 log quanta) per cone per second, the required energy would be disproportionately large and the curves above this point would show an increasing slope. However, they remain accurately proportional, implying full linearity.

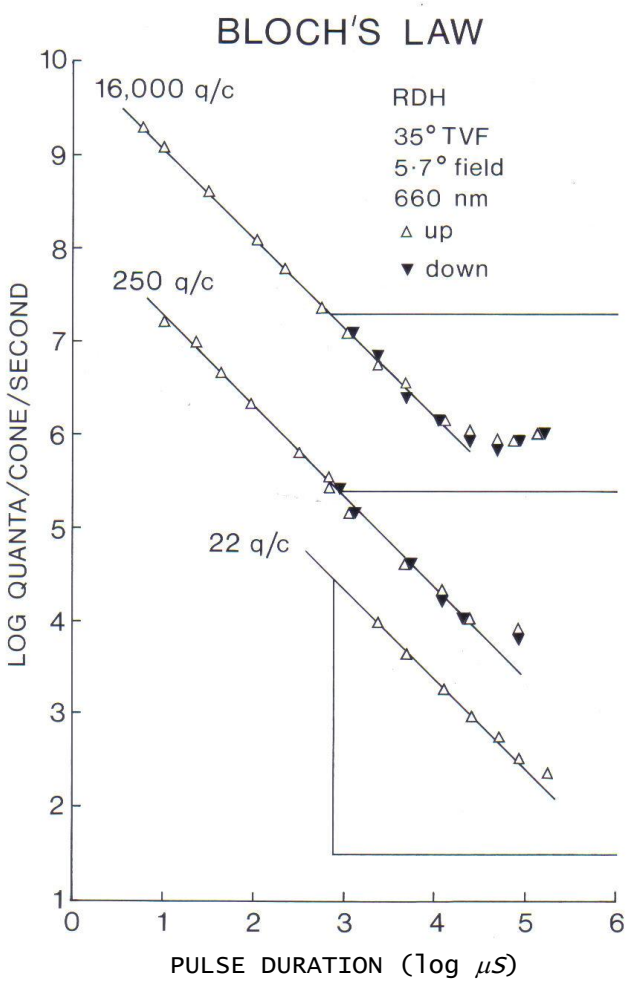

Fig. 5. Bloch's Law linearity and the temporal integration limit across the intensity range. Psychophysical thresholds were measured with an optimized staircase method for high-intensity 660 nm light-emitting diodes projecting to a 5.7° field at 35° eccentricity on the horizontal temporal meridian. The stimuli had base luminances of 0.7, 4.7 and 6.7 log cd/m^2 (intensities of 1. 4, 5.4 and 7.4 log quanta/cone/s at this retinal location) with a 180° white surround of matching intensity. Open symbols: pulse increments; filled symbols: pulse decrements. Horizontal lines: base intensities for each dataset

The second linearity test is to compare detection thresholds for pulse *increments* and pulse *decrements* (open and filled symbols, respectively, in Fig. 5). This comparison is possible only for threshold amplitudes less than the base luminance level because the luminance decrement cannot exceed its base level. Where the comparison could be made, there was no significant difference between the increment and decrement thresholds, implying that the temporal response of the human visual processing is linear with respect to the sign of the internal response.

The final issue addressed by these psychophysical data is that temporal integration fails beyond the range of the neural impulse response, and the detection threshold conforms to a horizontal line (Regan and Tyler, 1970; Gorea and Tyler, 1986; Sekiguchi, Williams and Brainard, 1993). This slope transition occurs at about 15, 20 and 100 *ms* for the three stimulus intensities (22q/c, 220q/c and 16000q/c respectively), which are typical values for this eccentricity. Since the highest intensity is close to the limit that is comfortable for the human viewer, and the lowest is close to the cone detection threshold (a 6 log unit range), this range of transition durations corresponds to the range of neural integration times expected for human vision. Thus, these psychophysical measurements provide an important constraint on the range of integration times operative in the underlying neural signals. These in turn constrain the neural response time constants allowable for the "temporal microscope" analysis of the neural response properties driving human BOLD signals.

In this connection, it should be noted that the temporal properties of the human psychophysics are compatible with the known properties of the responses recorded at the outer segments of human cones. Recordings from monkey cone outer segments (Schneeweis and Schnapf, 1995, 1999) show voltage response with roughly the right properties to account for human psychophysics, showing early response peaks with time constants of the order to 20 *ms*. The recordings in Schneeweis and Schnapf, (1995) match those from human cones (Kraft, Schneeweis and Schnapf, 1993), implying that voltage responses for human and monkey cones would also be similar. The difference between that 20 *ms* estimate and the 15 *ms* of Fig. 5 can easily be accounted for by the difficulty of recording up to the highest intensities, and the use of filter cutoffs no higher than 100 Hz in the physiological recordings.

Stepwise Incremental Estimation of the Neural Dynamics

One useful approach to the issue of the neural dynamics underlying the BOLD response has been published by Bandettini and Ungerleider (2001), on the basis of a temporal summation experiment reported in Birn, Saad and Bandettini (2001). The stimuli consisted of counterphasing checkerboard patches with durations increasing in factors of 2 from 0. 25 to 2 *s* (together with a 20 *s* duration). The key result was that the BOLD responses did not increase proportionately with stimulus duration, as would be expected for linear summation of a neural response of constant amplitude through convolution with the BOLD impulse response kernel. Instead, the amplitude showed little increase from the short durations, implying either a nonlinear summation process or a disproportionate strength of neural signals for short duration stimuli (Fig. 6).

Bandettini and Ungerleider (2001) proposed an analysis for the derivation of the implied neural signal under the assumption that the integration process underlying the generation of the BOLD response was indeed equivalent to a linear convolution operation. To derive the neural signal estimate, they took the difference between BOLD response areas for each successive pair of stimulus durations as an estimate of the amplitude of the neural signal for the time interval defined by the difference in stimulus durations. The result is reproduced in Fig. 6***b***. It implies that the neural signal has a pronounced early transient followed by a sustained phase of stable amplitude out to 20 s (red box function).

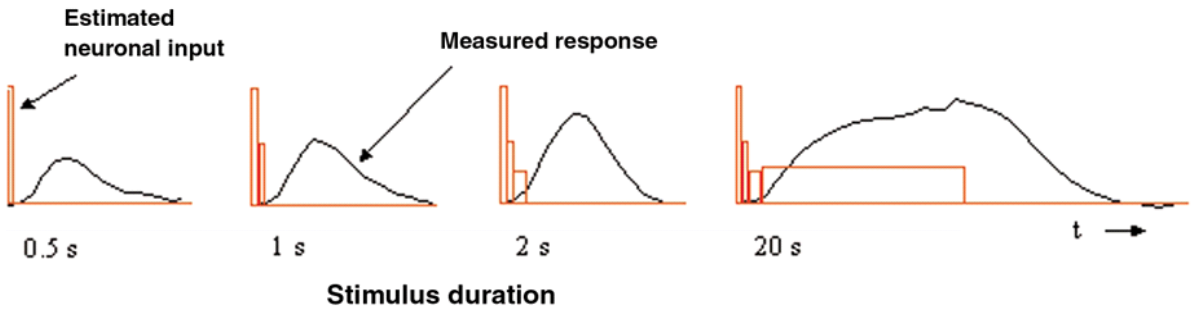

Fig. 6. Estimated neuronal input by a (simplified) stepwise forward model (Bandettini & Ungerleider, 2001) of the underlying neuronal response. This procedure derives the incremental neuronal signal from each stimulus duration (red box functions) as a uniform amplitude set by the amount of area that the BOLD response (black curves) increased over that for the next shortest stimulus. The result is a stepwise approximation to the underlying neuronal signal, showing that it has an early transient. This simplified stepwise model, however, is limited to the temporal resolution of the steps, of which the shortest is 500 ms.

As revealing as this analysis was, however, it does not seem to have been followed up by further estimates of the neural signals underlying the BOLD response to other stimulus types or in other cortical regions. Perhaps the reason is that the analysis requires an extended series of stimulus conditions. It should be clarified that the 250 *ms* neural transient estimated from the Bandettini and Ungerleider (2001) analysis is resolved only to the duration specified by that stimulus condition. The true underlying neural transient could have been much shorter. For example, the analysis could not distinguish between a neural transient of 250 *ms* and one of the 40 *ms* duration typical of neural responses. Only if the stimulus series had the requisite temporal sampling could the actual duration of the neural transient be determined (under the linearity assumptions of the analysis). Birn, Saad and Bandettini (2001) studied 5 durations, but a series of 12 or more would be required to cover the required range expected for neural signals (even under their assumption that the signal fluctuations would be likely to be proportionately slower as duration increased, permitting the temporal sampling to be logarithmically spaced). Thus, the neural integration time is of the order of 10 *ms* (Fig. 5), so stimulus durations of, say, 8, 16, 32, 64, 125, 250, 500 *ms*, and 1, 2, 4, 8, and 16 *s*, would adequately cover the required range with log-spaced sampling. If there were reason to suspect transient but delayed neural signals to be included in the response of a particular cortical area, however, a finer sampling density would be required. BOLD

responses for stimuli of very short duration have recently been reported by Uludag, Yesilyurt and Ugurbil (2006), so this approach to neural signal estimation should be practicable, even though intensive stimulation conditions would be required to obtain adequate sampling resolution.

Dynamic Forward Modeling

The analysis of Fig. 6 is a rudimentary kind of modeling, in the sense that it begins with an assumed form for the underlying neural signal. This form is a boxcar, based on the (unstated) assumption that the neural response is fast enough to follow the stimulus exactly, within the resolution visible in the plotted functions. The neural response is then convolved with a metabolic response kernel (***MRK***(t) in Fig. 1) to generate a model BOLD response that is fitted to the data with the amplitude as one free parameter. This parameter then, according to the linearity assumption, determines the inferred amplitude of the neural signal (red box function in Fig. 6***b***) that must have given rise to the measured BOLD response. This logic was applied by Bandettini and Ungerlieder (2001) in incremental fashion to each new temporal segment of the stimulus.

We can expand the forward modeling approach to reveal some remarkable properties of the neural response to extremely brief stimuli such as those reported by Uludag, Yesilyurt and Ugurbil (2006). By the use of high luminance LED stimulators, this group have been able to record BOLD responses to stimuli as brief as 1 *ms* (Fig. 7***a***). This achievement does not imply, of course, that the underlying neural response is as brief as this, but their data provides the opportunity to apply the forward modeling concept to a model neural signal that has full temporal resolution. We will show that the *forward modeling* approach is able to reveal precise details of the neural response at its native time resolution. Based on an appropriate model of the temporal summation experiment, there is essentially no limit to the temporal resolution to which the underlying neural response can be determined.

The starting point for this analysis is the model neural response function we have developed to fit the LFP data of Fig. 3. This function was convolved first with the stimulus time-course and then with a unitary MRK to generate a predicted cortical BOLD summation function (Fig. 7***b,*** blue dashed curve). The form of the predicted response was then optimized for best fit to the actual BOLD responses of Fig. 7**a** by varying the parameters of the model neural response (which operates like a hidden layer in a neural network model). The integrated areas for the optimized BOLD fits are plotted in Fig. 7**b** for comparison with the empirical area data. The resultant estimate of the neural signals are shown at two time scales in Fig. 7**c,d**. This analysis again implies that there is a neural transient followed by an extended sustained plateau, but now reveals the precise temporal properties of the neural response. The transient peaks at about 20 *ms* and that the plateau has an adaptive decay with a time constant of about 800 *ms*. Thus, the *forward modeling* approach in combination with the temporal integration paradigm can overcome the inherent limitations of the slow BOLD signal to quantify details of neural processing at the millisecond time scale.

Note that a single exponential attention term can account well for the integrated response growth beyond 500 *ms*. However, below this duration, no adjustment of the

relative weightings could improve the fit to the temporal integration data. Running the optimization with the two-component function of Fig. 4**c** results in the neural response function (red curves in Fig. 7***c,d)***. The fit is improved by a factor of three (which is a significant improvement on the chi-square test at $p < 0.01$). Note that resulting neural transient implied by this optimization to the BOLD signals is so brief that the time axis had to be scaled to 60 *ms* to show its features (Fig. 7***d***).

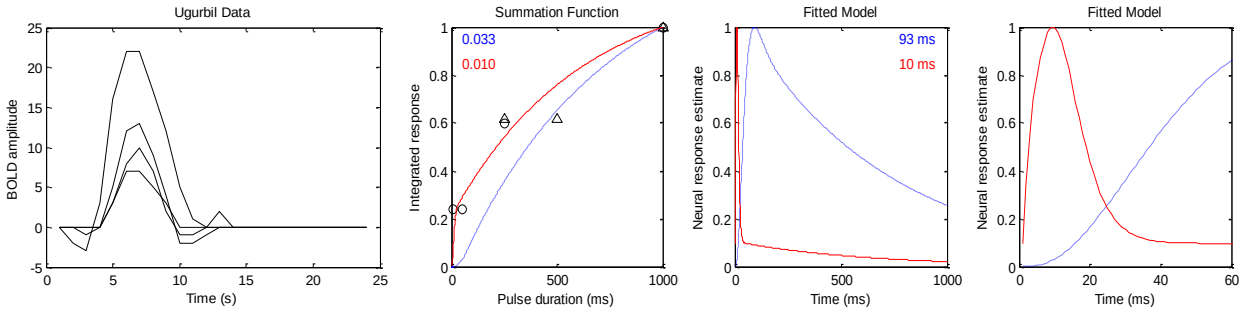

Fig. 7 **a**. BOLD responses to high-luminance flashes from for stimulus durations of 1, 3, 5, and 1000 *ms* duration (curves of increasing amplitude), replotted from Uludag et al. (2006). ***b***. Integrated BOLD response data from ***a*** (open circles), together with further areal integration data from Birn et al (2001, open triangles), for stimulus durations of 5, 50, 250 and 1000 *ms*. Dashed blue curve shows the initial predictions of the neural-response model prior to optimization. The full red curve shows the *optimization* of this model to the BOLD data depicted. Blue and red numbers represent the residual errors for the two cases. ***c***. Non-optimized (blue) and best-fitting (red) neural response estimates on a long time-scale. Blue and red numbers represent the peak latencies for the two cases. ***d***. The early phase of the non-optimized and best-fitting neural responses on a short time-scale. This result illustrates that forward modeling of temporal summation data can resolve details of the underlying neural response in human cortex on a millisecond time scale.

This characterization of the *neural* response characteristics in the human brain makes it clear that the *dynamic forward analysis* is capable of resolving differences in neural response time constants of the order of a few milliseconds. Previous studies have established that fMRI can resolve neural time delays of the order of 100 *ms* or less (Menon, Luknowsky and Gati, 1998). However, this range still falls short of resolving temporal properties in the typical range of neural transients. The present analysis shows that, with the right paradigm, the finest details at the level of neural temporal processing are accessible from analysis of the fMRI signal. We emphasize, of course, that this analysis is applicable to every voxel of fMRI recording, and thus the responses of local neural populations can be resolved to the millisecond level at a spatial resolution of the order single hypercolumns anywhere in the cerebral cortex.

Nonlinear Dynamic Forward Modeling

In this section, we propose a more biophysically-based development of the *forward modeling* approach that we term *Nonlinear Dynamic Forward modeling* (NDF modeling). Rather than simply characterizing the nonlinearities in the behavior of the BOLD

waveform (Birn, Saad and Bandettini, 2001; Bandettini et al., 2002) or attempting to infer the potentially complex properties of the underlying neural mechanisms from the form of the BOLD response by deconvolution (Glover, 1999; Logothetis, 2002, 2003a,b; Logothetis and Wandell, 2004), the concept of forward modeling is to start *from the 'cause'* – the neural signal, instead from the 'consequence' – the BOLD signal. Thus we incorporate into the model as much knowledge as possible about the likely neural substrate and then optimize the details to best fit the BOLD waveform. This knowledge includes known nonlinearities both of the population of neural responses to the stimulus but also of the metabolic requirements of the neural processing that lead to the measurable BOLD response.

The *dynamic forward modeling* discussed in the previous section is *linear* in the sense that it assumes no nonlinearity in the model linking the neural response to the BOLD response (or indeed in linking the stimulus waveform to the BOLD response) [1]. It merely estimates the *gain* of the neural response for each stimulus duration, which can vary by a linear process. For example, the transient neural response to a contrast onset would have such a characteristic, as detailed in the previous section.

A more general form of forward modeling is to incorporate a range of possible nonlinearities into the structure of the model, which we term *nonlinear dynamic forward* (NDF) modeling. The range of possible nonlinearities at each stage of the model is vast, so the approach is not feasible unless one incorporates Bayesian constraints in the modeling, based on the known biophysics of the neuronal responses likely to underlie the BOLD activation. The groundwork for such constraints has been provided by Logothetis and Wandell (2004). Thus, the starting point for the NDF modeling will be the neural signal, whose first effect in terms of the chain of BOLD dynamics is to create a metabolic demand *G(t),* in the neighboring glial cells *(*see Fig. 1).

Since little is known about the glial dynamics of transmitter recovery, we will pursue two options as to their effects. One is that the *metabolic demand* derives from the transmitter recovery cycle following the release of transmitter consequent to an axonal spike. Since axonal spikes represent only the positive aspect of the intracellular voltage and since 90% of cortical synapses are excitatory (Shank and Aprison, 1979; Wang and Floor, 1994), the signal transmitted from one cortical stage to the next may be treated as a half-wave rectified version of the dynamic neural signal. This prediction is shown as the blue curves in Fig. 8***Aa***, which is an overlay of the model estimates of the neural responses to stimulus pulses that double in duration from 8 *ms* to 16 s (eight doublings). For this example, the neural response has balanced excitation and inhibition, so there is only an initial transient response even to the prolonged pulses, with the negative lobe at offset being thresholded out by the rectification. (Note that the local metabolic demand, M_i(t), per se, has the same time course in this model as the transmitter recovery from

[1] Strictly speaking, many GLM models contain an implied nonlinearity, in that a flickering counterphase stimulus should generate both positive and negative activations in a typical neural array, tending to cancel to a null signal. It is only if there is a rectifying nonlinearity in the neural response that they should sum to a waveform approximating the integrated amplitude of the stimulus waveform. However, we are adopting the common convention here that a neural signal matching the net stimulus contrast over time is considered to be linear.

which it derives. The energetic processes required for the recovery of the initial state, however, form an oxygen-based chain of glial metabolic response, *G*(t), that may have substantially slower time course at one or more stages.)

The other option is to consider the instantaneous metabolic demand of both polarities relative to the mean background level (i.e., the demand of both excitatory and inhibitory cells, or both 'on' and 'off' cells), implying that the signal generating the metabolic demand is a full-wave rectified version of the intracellular voltage. This option is shown in Fig. 8***Aa***, as the red curve, whose spikes represent the instantaneous metabolic demand for the offset transient for each of the pulse duration doublings (which were thresholded out in the blue curve).

We may now pursue the response to these two options for the nonlinearity of the metabolic demand through the biophysical chain to the measured BOLD signal. The first element in this chain is the astrocytes, which provide glucose to the neuron and replenish its supply by ATP metabolism fueled by oxygen from neighboring blood vessels. It is important to emphasize that the astrocyte metabolic processes are as slow, relative to the intracellular signal dynamics, as are the processes of hemodynamic oxygen supply. The time constant of the astrocyte responses is known to be of the order of several seconds (Kelly and van Essen, 1974; Filosa, Bonev and Nelson, 2004; Metea and Newman, 2006), and it is clear that there must be a substantial pre-hemodynamic component from these slow responses. However, at present too little is known of their dynamics and/or nonlinearities to securely assign precise time constants to the astrocytic component relative to the hemodynamic component. We will therefore treat the entire chain from the metabolic demand to the magnetic resonance signal in the traditional fashion, as a unitary linear kernel. This kernel is often termed the hemodynamic response function (HRF), but in view of its likely substantial astrocyte contribution, we will give it the more general term of the metabolic response kernel (MRK).[2] Our main goal is to estimate the properties of the neural signal processing, and it will be seen that there is sufficient information to provide a rich analysis of these properties, and to account for the empirical nonlinearities of the BOLD signal, as long as the metabolic supply chain conforms to the linearity assumption.

The way in which NDF modeling captures the underlying neural properties is laid out in Fig. 8 (indexed in row/column notation). This figure shows the consequences of the assumption that the neural impulse response has balanced positive and negative lobes, and of various degrees of imbalance of the negative lobe. It also compares two forms of nonlinearity of the metabolic demand. (The figure does not show the neural responses themselves, but takes the metabolic demand as the starting point for the analysis.) The point of the figure is to illustrate that subtle properties of the neural response and its metabolic effects have dramatic effects on the measured BOLD waveform, and that therefore the BOLD waveform characteristics in well-controlled experiments can be diagnostic of the fine details of the neural and metabolic processes in the brain regions from which the BOLD signals derive.

[2] The term "kernel" is preferred because it is conceptualized as the theoretical response to an imaginary stimulus of infinitely short duration and unit area, termed a "Dirac pulse".

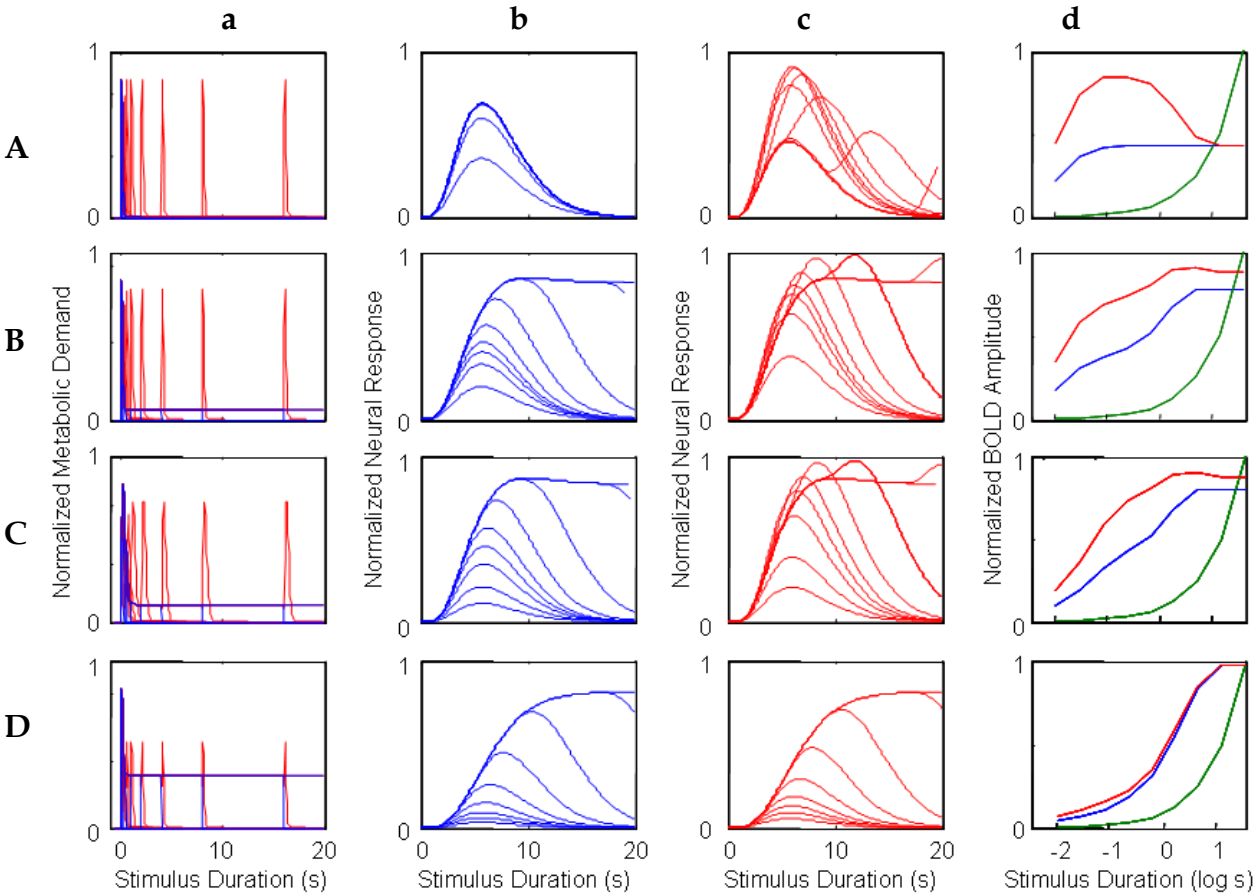

Fig. 8. Simulations of four different types of BOLD response for monophasic metabolic demand signals (and a monophasic Metabolic Response Kernel). The rows represent the results for **A**: metabolic demands with a purely transient time course, **B-D**: responses with a mixed transient and sustained time course, in which the sustained component is respectively 12%, 18% and 50% of the amplitude of the transient component (based on different ratios of neural excitation/inhibition). For each type, column (***a***) shows the assumed metabolic demand, column (***b***) plots the BOLD responses over duration for the half-wave-rectified model of metabolic demand, column (***c***) plots the BOLD responses over duration for a full-wave-rectified model, and column (***d***) plots the duration summation curves for peak amplitude (blue curve: half-wave-rectified model, red curve: full-wave-rectified model, green curve: full linear summation). Note the use of the logarithmic abscissa to focus on the brief duration regime. The progression of the model BOLD responses and the form of the summation curves are diagnostic of both the relative weighting between the sustained and transient components of the neural signal and the form of rectification feeding the metabolic demand.

The first row of Fig. 8 shows the development of the BOLD response waveform through the pulse duration series when the metabolic demand derives from a balanced neural signal. Half- and full-rectified versions of the metabolic demand are shown in Fig. 8Aa (blue and red curves**).** The resultant BOLD waveforms are computed under the assumption of a gamma function model of the MRK (HRF), as developed in Fig. 2. Because the balanced neural signal generates only a spike in the half-rectified version

(column ***b***) of the neural metabolic demand, the predicted BOLD waveform approximates the form of the MRK for all durations, varying only in amplitude as the neural signals integrate the energy in the pulse stimuli. After the first two durations the BOLD amplitude asymptotes to a stable value.

The development of the peak amplitude in the BOLD temporal summation series is shown as the blue summation curve in Fig. 8***Ad***. The critical point of this plot is that the asymptotic corner of this summation curve occurs at 40 *ms*, which is the value of the time constant assumed for the neural signal. Thus, the form of the BOLD amplitude summation series (Fig. 8, column ***d***) is diagnostic of the time constant of neural integration down to the millisecond range. There is no limit in principle to the temporal resolution that can be achieved by this methodology since it is estimated from the amplitude variation of the BOLD signal as a function of stimulus duration, not from the its temporal aspects.

The reduction in peak amplitude is captured in the red temporal summation curve of Fig. 8***Ad***, showing a reduction by a factor of two for long durations. (The green curve in Fig. 8***Ad*** represents the values expected for fully proportional summation of the energy in the stimulus pulse; it is an accelerating curve due to the logarithmic abscissa.)

It is evident that the temporal summation properties of this first model of the nonlinear dynamics is not close to those observed in typical fMRI studies (such as those reproduced in Fig. 7), but it is presented to illustrate the power of NDF modeling to evaluate particular properties of the neural signal dynamics. In this case, the main culprit is the exact balance of excitation and inhibition in the model, bringing the response for long-duration stimuli back to baseline after about 15 sec (when typical data show a much slower decline over time). However, notice that relaxing the assumption of such a balanced response by even a small amount is enough to obtain much more plausible predictions.

The case of the full-rectification model of the metabolic demand is shown in Fig. 8***Ac.*** Here the second neural response peak (i.e., that from the stimulus offset) plays key a role in varying the BOLD waveform, which first extends in time and then shows a two-peaked structure with reduced amplitude for long-duration stimuli.

Fig. 8***Ba*** shows the *half-* and *fully-rectified* version of the metabolic demand to the same pulse duration series, where the neural inhibition is now assumed to be reduced in energy by 1.5% relative to the excitation. This small imbalance is magnified by the convolution with the sustained stimulus, and thus it results in a sustained component that is 12% of the amplitude of the initial transient (blue curve in Fig. 8***Ba***) and then into an almost fully sustained set of BOLD response functions (Fig. 8***Bb***, compare to ***Ab***). Thus, the *form* of the BOLD response functions can be strongly *diagnostic* of even slight variations in the properties of brief *neural* signals. Again, the nature of the metabolic demand function (half- or fully-rectified) has a big impact on the form of the BOLD response, determining whether an offset peak is evident at the tail of the responses even when they are sustained (Fig. 8 ***Bc***). Such a peak has been reported in some studies but is not always evident. Thus it remains an empirical question to what extent it is

representative of BOLD waveforms. Intermediate forms of the rectification model are required to capture the empirical properties in detail.

Note that the amplitude series in Fig. 8***Bb*** and **c** show bands of denser packing, where the amplitude changes were not proportional to the doublings of stimulus duration. This property has often been called a "nonlinearity" of the BOLD response, but it is generated here by a linear convolution process, so for clarity we will refer to it as a "non-proportionality" (see Section 3). Viewed in terms of the sequence of BOLD waveforms in Fig. 8***Bb,c***, the regions of dense packing form an intermediate "shelf" or partial asymptote in peak amplitude summation plots of Fig. 8***Bd***. It is again evident that the onset of this intermediate 'shelf' in the summation curve corresponds to the integration time of the underlying neural signal, while the second asymptote corresponds to the 5 sec integration time of the MRK (HRF). Careful measurement of such summation functions can therefore provide estimates of both the neural and the metabolic time constants in the neural > BOLD signal chain.

This point is emphasized by the response set in row ***C*** of Fig. 8. The key difference from the parameters used in row B of Fig. 8, is that the neural time constant was doubled from 50 *ms* to 100 *ms* (and the excitation/inhibition imbalance was also increased to 7% to maintain the same form of offset peak). It is evident that (i) the summation curve (Fig. 8***Cd***) takes a measurably different form, and that (ii) the accuracy of estimation of the neural time constant is limited not by the BOLD time constant but only by the variability of the BOLD amplitude measures. For example, this analysis shows that the neural time constant is estimable to an accuracy of about 0.=1 log units if the BOLD response functions can be measured to an achievable accuracy of about 10%. Note that the values of 50 and 100 *ms* were chosen here for illustrative convenience, and the estimation of a time constant of, say, 10 *ms* should be equally good.

The final case (Fig. 8***Dd*)** shows the NDF predictions of increasing the excitation/inhibition imbalance of the neural response to 50% in order to model a system that is predominantly sustained in nature. Under these conditions, the impact of the initial transient becomes essentially negligible, and the summation curves (Fig. 8***Dd)*** become indistinguishable from proportional summation (i.e., they run parallel to the green curve). This manipulation illustrates that the power of the NDF analysis depends on the neural processing being predominantly transient, and that the properties of the underlying neural mechanisms would not be accessible to this form of analysis in predominantly sustained systems. Luckily, however, the well-established deviation from proportionality for short-duration stimuli implies that the neural system is, in practice, predominantly transient and is therefore amenable to this form of analysis.

In summary, NDF modeling of the full-range duration summation series provides a powerful tool for the analysis of detailed properties of the underlying *neural* signals, including their *time constants*, their *excitation/inhihibition ratio* and the form of *nonlinearity* coupling them to the metabolic recovery processes.

Multicomponent Optimization of the NDF Model to BOLD Responses for a Family of Luminance-Based Stimuli

Background

We have empirically evaluated the NDF model on the basis of BOLD fMRI data that we have recorded in visual stimulation experiments in human. Critically, we have estimated the neural signals underlying the BOLD responses generated by different stimuli *in one and the same* ROI. The rationale behind this study is the fact that BOLD waveforms within one and the same cortical region may vary dramatically as a function of the stimulus type, in spite that the metabolic/hemodynamic response functions (MRK) must remain invariant in the region (Likova and Tyler, 2007; d'Avossa, Schulman and Corbetta, 2003). Consequently, differences in the BOLD response waveform must be interpreted as due to differences in the neural response waveforms to the different stimuli.

The stimulus paradigm that we have previously developed following this logic generates distinct percepts of spatial structures from different but very brief. The BOLD response to any impulse0like stimulus is traditionally considered to have a constant shape and is used as the HRF estimate in the most popular methods and software packages (*e.g .*, SPM, etc.). However, measuring the waveforms for a whole set of different 'impulse' stimuli revealed significant waveform deviations within *the same* cortical areas, thus ensuring that they reflect neural processing differences specific for each of the visual stimuli.

Experimental paradigm

For the purposes of the empirical evaluation of NDF modeling, we have developed a stimulus paradigm including a set of four basic types of luminance transients. They were based either on a uniform field or a bar-grating of 6 arcmin squares, designed to target mechanisms with either low or high spatial-frequency tunings. The uniform gray field was switched to either zero or double the gray luminance for a pulse of 400 *ms* duration. The grating stimuli had 100% luminance contrast and were presented from a uniform gray pre-stimulus state for either 100 *ms* or 400 *ms*.

Results

An example of the variety of cortical responses obtained in retinotopic area V1 combined for both hemispheres (blue curves) is shown in Fig. 9, row ***A***, together with the monophasic MRK (green curves) estimated for the same subject. For the grating stimuli (upper panels), the response was almost monophasic for the 100 *ms* duration but was strongly biphasic for the 400 *ms* duration. For the dark pulse, the response was again approximately monophasic, while the response to the bright luminance pulse shows a striking increase in the BOLD signal many seconds after the primary peak.

A similar variety of BOLD response types was seen in the response from cortical area hV4 (Fig. 9, row ***B***), although here the pattern was different from that for the early visual area. The responses to the grating stimuli different from those for V1 with a

stronger early peak for the 100 *ms* duration that disappeared for the 400 *ms* duration, and was replaced by a deep negative response. However, the response for the dark and bright pulses were similar in the two areas, showing and enhanced positive response relative to any of the other stimulus types, while that for the bright pulse had both early and late peaks of entirely positive response. Thus, for both cortical regions, the variety of response waveforms implied a large variation in the neural signals for the four brief stimulus types (whose duration was short enough that they would each be expected to generate the same PERF under the standard GLM analysis; green dashed curves in Fig. 9.)

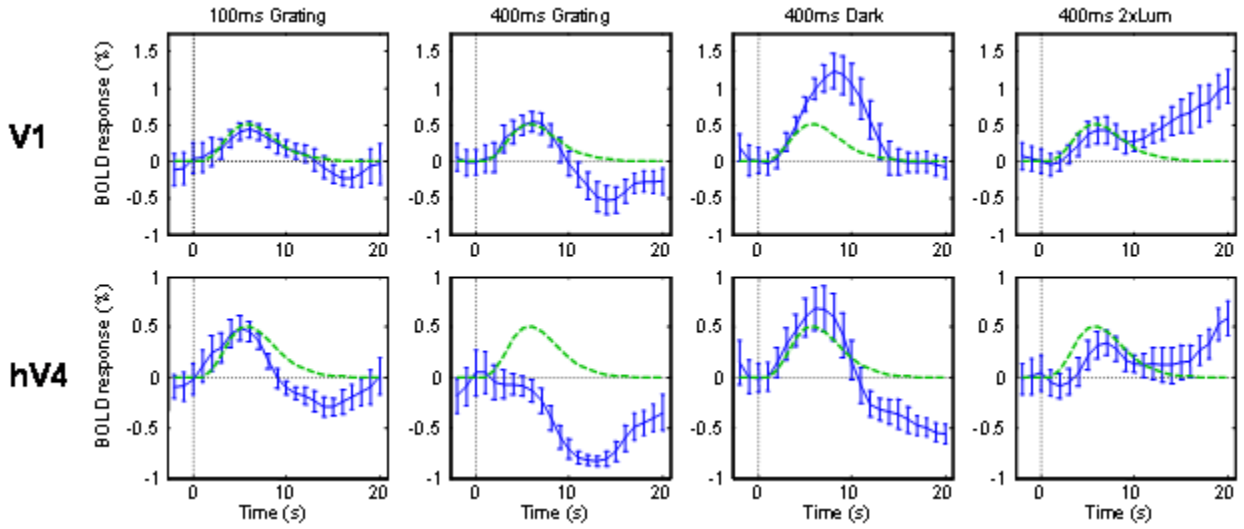

Fig. 9. Variety of BOLD response waveforms obtained in retinotopic areas V1 and hV4 for the set of brief grating stimuli of 100 *ms* and 400 *ms* duration, and uniform-field increment and decrement stimuli of 400 *ms*. Note that many of these responses differ substantially from the form of the typical MRK (dashed green lines). The BOLD waveforms differ within each retinotopic area, although the similarity of characteristics between the two areas indicates that the differences are stable and repeatable.

Analysis

Our analysis utilizes the *Nonlinear Dynamic Forward* (NDF) modeling approach based on the assumptions of Fig. 1, that the neural signals from the neural populations within a small region of cortex generate local metabolic demands that may be nonlinearly related to the neural signals but which sum linearly to form the net activation signal. The linear convolution of the net activation signal with a metabolic response kernel (MRK) thus generates an adequate prediction of the recorded BOLD signal. The nonlinearity lies in the presence of multiple summed components, each derived by linear convolution of the component metabolic demands with the unitary MRK. It may also incorporate the rectification process illustrated in Fig. 8 (red curves), but such a rectification was not needed for the present analysis.

To analyze the BOLD response waveforms, we assume a blend of two types of gamma functions, based on common features of responses recorded from neurons in

early visual areas, namely with exponent values of k = 1 and 5. These values generate, respectively, a brief and prolonged form of delayed gamma function. These two components are designed to constitute a minimum set of the neural signal properties required to account for the varieties of BOLD responses that are encountered in Fig. 9. The components were optimized to the waveform for each stimulus type simultaneously with the optimization of a unitary HRF (Fig. 10*A*) for the transmutation of the metabolic demand into a predicted BOLD waveform. Note that the HRF used for this analysis is itself assumed to be from the family of (monophasic) gamma functions, on the grounds that many BOLD waveforms are found to be monophasic. The monophasic form of many empirical BOLD waveforms tends to exclude the possibility that the HRF is, in general, multiphasic because it is highly implausible that the neural signal underlying a monophasic BOLD response would happen to have the right form to cancel the multiple phases if the HRF was, in fact, multiphasic.

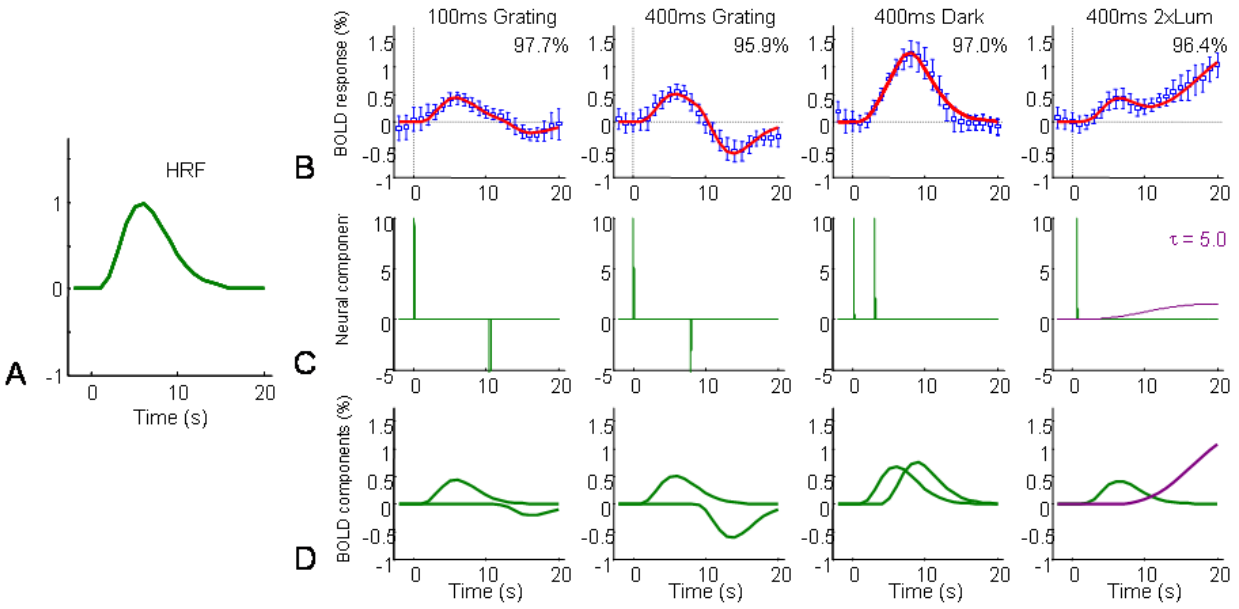

Fig. 10. Dynamic forward modeling optimization of neural components to the BOLD responses for cortical area V1, with simultaneous optimization of the unitary HRF to convert the neural signals into the BOLD waveform predictions. ***A***. Optimized monophasic HRF. From left to right, the data in the columns correspond to (i) 100 *ms* grating, (ii) 100 *ms* grating, (iii) 400 *ms* dark pulse (decrement), (iv) 400 *ms* bright pulse (increment). ***B***. Measured BOLD waveforms for each stimulus type (blue) with best-fitting waveforms (red) and percent variance accounted for. ***C***. Optimized neural components for the two-component case. ***D***. Individual BOLD waveforms for the separate components making up the summed BOLD waveforms in ***B***. Two-component optimizations gave satisfactory fits to all waveforms with minimal delay in the first neural components. Different temporal properties are selected by the optimization to account for the variety of different waveforms.

The novel analysis we propose consists of (i) the simultaneous optimization of the a set of neural components (ii) convolved with the unitary HRF separately for all four stimulus types together with (iii) the rise and fall parameters of HRF. The results of the analysis for the V1 dataset for the battery of four stimulus types we have run with fMRI, is that the two components were sufficient to account for >95% of the BOLD variance in each case. Although the average responses exhibit a wide variety of waveforms, the unitary HRF (Fig. 10***A***) has the typical form known from previous studies, peaking at about 6 s and returning to baseline by about 18 s. The center row (Fig. 10***C***) shows the optimized pair of neural signals in each case, with their parameters optimized to provide the fits shown in the top row (Fig. 10***B***). The contribution of each component to the BOLD signal is shown in Fig. 10***D***. Note that, although the delay parameter was free to vary, the delays remained close to zero for each primary positive component. The second component was negative (inhibitory) in two cases, with either long delays or a very slow time course, illustrating the unexpected properties of the components required to account for the full variety of BOLD waveforms. A similar degree of success was obtained for other datasets, including the hV4 data of Fig. 9B.

Of course, any such estimation of the parameters of a set of components underlying a noisy empirical function will have some error variance. A standard approach to the estimation of the parameter variance is the jackknife statistic (Pitman, 1937; Miller, 1964). In general this procedure consists of the removal of each measured point of a function and successive refitting of the parameter estimation to the remaining data for each "*k*th deleted" dataset. The variance of the fits to the loss of each measured point can then be used to infer the overall variance. This approach provides a means of determining the sensitivity of each parameter to the noise contaminating the measured responses. Thus, the jackknife approach provides a means of setting the confidence interval on any of the measured parameters. This capability will be particularly relevant in the next section of the application to the measurement of the causal dynamics of information flow through the cortex.

Causal Cortical Dynamics

One of the key aspects of neural processing that needs to be determined is the flow of information among cortical and subcortical areas. When two or more brain regions are simultaneously active, the direction of influence among between these areas is ambiguous. The effective connectivity approach (Friston et al., 1993a; Friston, 2001) has been proposed to resolve the direction-of-influence ambiguity by defining explicit statistical models of directed neuronal interactions, and it was further advanced by covariance structural equation modeling (Buchel and Friston, 1997; McIntosh and Gonzalez-Lima, 1994), nonlinear system identification techniques (Friston and Buchel, 2000), such as Bayesian estimation of deterministic state-space models (Friston et al., 2003). However, the classic 'effective connectivity' methods have the problem that they

rely on a priori specification of a model that contains pre-selected regions and prior knowledge about the existence and direction of influence of the connections between them. Thus, a conceptual advance toward effective connectivity mapping without pre-selection was the use of "Granger causality" (Granger, 1969; Sims, 1980), to determine both the existence and the direction of interactions from empirical data. Granger causality analysis has been applied to electrophysiological animal data (Baccala and Sameshima, 2001; Bernasconi and Konig, 1999; Bernasconi et al., 2000; Brovelli et al., 2004; Freiwald et al., 1999) and EEG data (Hesse et al., 2003; Kaminski et al., 2001) and to fMRI data (Goebel et al., 2003, 2004). Consequently, various methods for modeling of fMRI time series in the context of Granger causality have been proposed, such as a vector autoregression (VAR) (Roebroeck et al., 2005) and wavelet-dynamic vector autoregression (Sato et al., 2006), among others. However, a key problem for the successful application of Granger causality to fMRI signals is its low temporal resolution, due to the fact that the BOLD signal is only an indirect measure of the neuronal responses and thus the neuronal responses are temporally blurred by the metabolic delays of this signal.

The NDF analysis of the underlying neural components sheds further light on the Granger causality. Granger (1969) argued that the flow of time is a natural property incorporated into our concept of causality, and that only associations with a finite lag should be considered to be causal. In its usual implementation, Granger causality is based on a cumulation of significant regression coefficients between two signals across time lags between the two signals. Only the regression coefficients with a time lag are incorporated in the Granger coefficient of causation to the lagged variable.

Thus, Granger causality is somewhat problematic when considered in terms of typical neural event sequences, where 'events' consist of signals with extended time courses. If one signal has a short time course and another has a long time course, for example, it is natural to think of their peaks as specifying the timings in their designation as events. If the signal with a short time course goes into a system with a long time constant it will 'cause' a prolonged 'event' peaking at a later time, conforming to the Granger concept. If, on the other hand, the signal with the long time constant has the nonlinear effect of triggering the signal with a short time constant at an early point in the rising phase (of the long-time constant signal), it could happen that the *caused* event peaks earlier than the causal event, reversing the order of inference in the Granger analysis. Thus, particular kinds of nonlinearity, such as trigger events, can violate the assumptions required by Granger, so its inferences should be regarded as probabilistic rather than definitive.

As stated, the Granger causality calculation has the implicit problem that it lacks temporal specificity because takes the *autoregressive spread* of information in each analyzed signal into account by cumulating over *all* delay intervals. The result is purely the causal direction and strength of the connection between any pair of signal sources. We propose an alternative approach based on *temporal cross-correlation* that provides *fine-grain temporal information* about the signal delays between the sources. When there

is delayed signal transmission from one signal 'channel' to another, the cross-correlation function between the two signals will exhibit an asymmetry with a peak at the lag corresponding to the transmission delay (Knapp and Carter, 1976). Cross-correlation captures the second-order moments of the full autoregression statistics of the Granger causality formalism. However, Granger causality loses the information about specific delays by summing the lagged mutual information over all lags. The *cross-correlation* approach allows the identification of the specific lag by the *autonormalization* of the two signals being compared in the cross-correlation to remove the effect of their temporal autocorrelations.

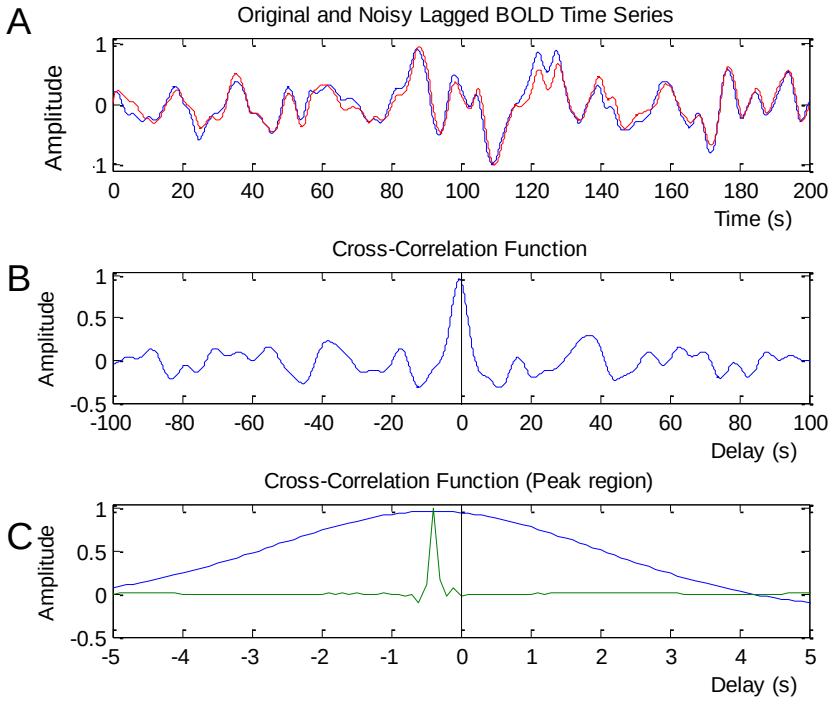

Fig. 11. Autonormalization of the cross-correlation for time-resolved delays. *A*). A random time series convolved with a physiologically plausible MRK giving it substantial self-correlations (blue) and a delayed noisy copy of the same signal (red). *B*). Cross-correlation between the original and delayed time series. ***C***). Expanded central region of ***B***(blue) with autonormalized version of the same function (green). Note that, for this case of a single delay, the autonormalization compresses the causal signal essentially to a single time sample, thus greatly increasing the temporal resolution of the method.

An example of a single lag between two signals is shown in Fig. 11. The broad spread of positive and negative lags of the original autocorrelation is converted to a narrow spike that clearly specifies the *unitary delay* characterizing the connection between the two sources. This compensation may be achieved simply by dividing the Fourier Transform of the cross-correlation signal by its Fourier Amplitude Spectrum, then transforming back to the time domain. This *autoregressive normalization* should account for the temporal spread of all signals feeding into each of the measured activation areas. The essence of the autoregressive*]* normalization is to collapse the spread of delays due to the build-up and decay in the responsive system to a single

characteristic time-shift representing the transmission delay between the two signals, which will be termed the *autonormalized delay*. From this perspective, each autonormalized delay represents a *specific causal link* between the two components. (As in all non-manipulative inferences of causality, there is always the possibility that an apparent causal link of A > B actually represents separate links of C > A and C > B, the latter having extra delay. Such possibilities can be excluded by direct manipulation of the signal at A. In the absence of such manipulation, the inference of causality can only be treated as the parsimonious hypothesis.)

If there is more than one autonormalized delay spike, it implies separate causal links between these systems. The usual implementation of Granger causality throws away this differential information by cumulating the (lagged) delays into a single index, but in the autonormalization approach each lag represents a specific causal link, allowing them to be considered separately rather than being cumulated together. Thus, our novel approach to causality treats each autonormalized delay as a separate causal link between pairs of signals in a network of interconnected cortical regions, with the ability to account for parallel causal links.

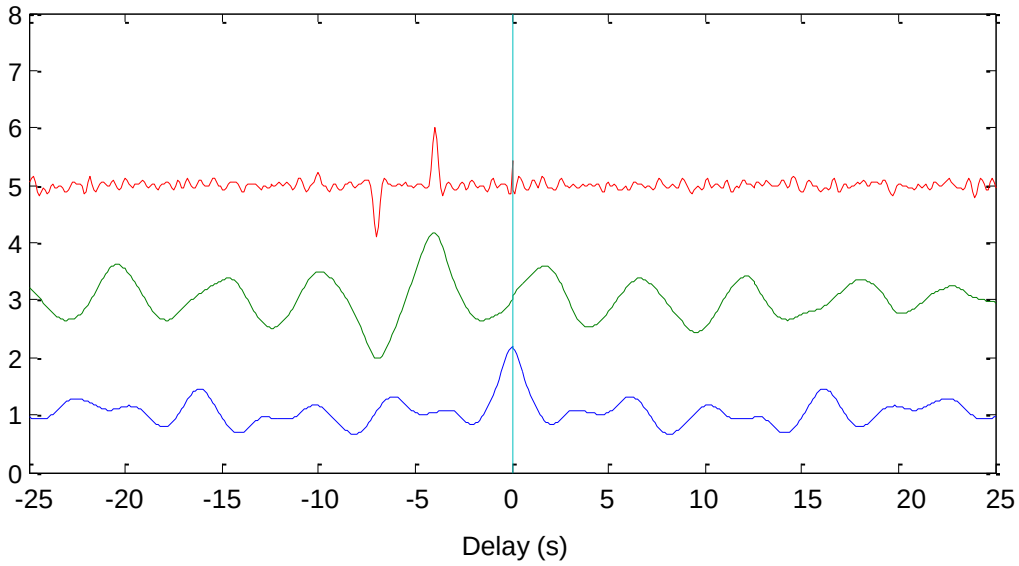

Fig. 12. Autonormalized analysis of inhibitory feedback. ACF of input signal (blue), CCF of output with (iterative) inhibitory feedback (green), and autonormalized CCF (red) revealing the excitatory (4 *s*) and inhibitory (7 *s*) delays.

The proposed autonormalization approach is also able of revealing an inhibitory influence between cortical regions, and indeed *separate excitatory and inhibitory* influences between the regions. A simulation of this capability is shown in Fig. 12. The simulated BOLD waveform was generated again as an MRK-filtered noise model signal that was delayed relative to its input signal and then fed back to the input with a negative weight. As a result, the autonormalized cross-correlation function between input and output (green trace) has *both* a positive and a negative correlation peak indicating delays of 4 s for the feedforward and 7 s for the feedback. These delays are

derived from the simulation based on realistic temporal properties of the BOLD signal, indicating that there is a far greater ability to resolve specific temporal interactions through the BOLD signal than previously appreciated.

Conclusion

The suite of techniques introduced in this paper can provide a significant enhancement in the ability to resolve the timing and neural signal estimation underlying the BOLD waveforms recorded throughout the cortex. Of course, the precision of the results is inherently limited by the signal/noise ratio of the recordings, but this precision can be enhanced by trade-offs against spatial resolution and by continued technological improvements such as multi-channel coils. The main point of this paper is to show that, given adequate signal/noise ratio, it is possible to develop approaches that overcome the temporal limitations of BOLD signal and are able to reveal the relevant properties of the underlying neural signals. These neural signal estimates may, in turn, be applied to the non-invasive analysis of deficits in the temporal processing of neurological diseases affecting cortical processing in the human brain. The neural delay estimates may, in turn, be applied to revealing the causal relations in understanding the whole-brain functional integration, and in particular to reveal the interactions in cortical networks of activated areas.

Acknowledgements

Supported by RO1 EY 7890 and the Pacific Vision Foundation.

References

Aubert A, Costalat R (2002) A model of the coupling between brain electrical activity, metabolism, and hemodynamics: Application to the interpretation of functional neuroimaging, NeuroImage **17**:1162–1181.

Baccala LA, Sameshima K (2001) Partial directed coherence: a new concept in neural structure determination. Biol. Cybern, **84**: 463-74.

Barlow HB (1958) Temporal and spatial summation in human vision at different background intensities. J Physiol, **141**:337-50

Bandettini PA, Ungerleider LG (2001) From neuron to BOLD: new connections. Nat Neurosci, **4**:864-6.

Barth M, Norris DG (2007) Very high-resolution three-dimensional functional MRI of the human visual cortex with elimination of large venous vessels. NMR Biomed. (in press).

Bellgowan PS, Saad ZS, Bandettini PA (2003) Understanding neural system dynamics through task modulation and measurement of functional MRI amplitude, latency, and width. Proc Natl Acad Sci USA, **10**:1415-9.

Bernasconi C, Konig P (1999) On the directionality of cortical interactions studied by structural analysis of electrophysiological recordings. Biol Cybern, **81**: 199-210.

Bernasconi C, von Stein A, Chiang C, Koenig P (2000) Bi-directional interactions between visual areas in the awake behaving cat. NeuroReport, **11**: 689-92.

Birn RM, Saad ZS, Bandettini PA (2001) Spatial heterogeneity of the nonlinear dynamics in the FMRI BOLD response. NeuroImage, **14**: 817-26.

Bloch AM (1885) Experiences sur vision. Comptes Rend Soc Biol, **2** : 493-495.

Brovelli A, Ding M, Ledberg A, Chen Y, Nakamura R, Bressler SL (2004) Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. Proc Natl Acad Sci USA, **101**:9849-54.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci, **16**:4207-21.

Buchel C, Friston KJ 1997) Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cereb Cortex, **7**:768-78.

Carandini M, Ferster D (1997) A tonic hyperpolarization underlying contrast adaptation in cat visual cortex. Science, **276**:949-52.

Carandini M, Ferster D (2000) Membrane potential and firing rate in cat primary visual cortex. J Neurosci, **20**:470-84.

Dale AM (1999) Optimal experimental design for event-related fMRI. Hum Brain Mapp, **8**:109-14.

Feigl H (1953) Notes on causality. In, Readings in the Philosophy of Science, Eds. H. Feigl and M. Brodbeck, Appleton-Century-Crofts, New York.

Filosa JA, Bonev AD, Nelson MT (2004) Calcium dynamics in cortical astrocytes and arterioles during neurovascular coupling. Circ Res, **95**: 73-81.

Freiwald WA, Valdes P, Bosch J, Biscay R, Jimenez JC, Rodriguez LM, Rodriguez V, Kreiter AK, Singer W (1999) Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex. J Neurosci Methods, **94**: 105-19.

Friston KJ (2001) Brain function, nonlinear coupling, and neuronal transients. Neuroscientist. **7**:406-18.

Friston KJ (2002) Bayesian estimation of dynamical systems: an application to fMRI. NeuroImage, **16**:513-30.

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci, **360**:815-36.

Friston KJ, Buchel C (2000) Attentional modulation of effective connectivity from V2 to V5/MT in humans. Proc Natl Acad Sci USA, **97**: 7591-6.

Friston KJ, Frith CD, Frackowiak RSJ (1993a) Time-dependent changes in effective connectivity measured with PET. Hum Brain Mapp, **1**: 69-79.

Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ (1993b) Functional connectivity: the principle component analysis of large (PET) data sets. J Cereb Blood Flow Metab, **13**: 5-14.

Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. NeuroImage, **19**:1273-302.

Glover G (1999) Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage. **9**:416-29.

Goebel R, Roebroeck A, Kim DS, Formisano E (2003) Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. Magn Reson Imaging, **21**:1251-61.

Goebel R, Roebroeck A, Kim DS, Formisano E (2004) Directed cortical interactions during dynamic sensory-motor mapping. In: Duncan J (ed.), Attention and Performance XX. Oxford University Press, New York, pp. 439-62.

Gorea A, Tyler CW (1986) New look at Bloch's law for contrast. J. Opt. Soc. Am. A3, 52-61.

Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods, Econometrica, **37**:424-438.

Handwerker DA, Ollinger JM, D'Esposito M (2004) Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage. **21**:1639-51.

Hansen KA, David SV, Gallant JL (2004) Parametric reverse correlation reveals spatial linearity of retinotopic human V1 BOLD response. NeuroImage. **23**: 233-41.

Harel N, Lin J, Moeller S, Ugurbil K, Yacoub E (2006) Combined imaging-histological study of cortical laminar specificity of fMRI signals. Neuro*i*mage. **29**:879-87.

Henson RN, Price CJ, Rugg MD, Turner R, Friston KJ (2002) Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. NeuroImage, **15**:83-97.

Hesse W, Moller E, Arnold M, Schack B (2003) The use of time variant EEG Granger causality for inspecting directed interdependencies of neural assemblies. J Neurosci Methods, **124**: 27-44.

Horwitz B (1990) Simulating functional interactions in the brain: a model for examining correlations between regional cerebral metabolic rates. Int J Biomed Comput, **26**: 149-70.

Horwitz B, Grady CL, Haxby JV, Ungerleider LG, Schapiro MB, Mishkin M, Rapoport SI (1992) Functional associations among human posterior extrastriate brain regions during object and spatial vision. J Cogn Neurosci, **4**: 311-22.

Kelly JP, Van Essen DC (1974) Cell structure and function in the visual cortex of the cat. J Physiol. **238**:515-47.

Kamiński M, Ding M, Truccolo WA, Bressler SL (2001) Evaluating causal relations in neural systems: granger causality, directed transfer function and statistical assessment of significance. Biol Cybern. **85**:145-57.

Knapp CK, Carter G (1976) The generalized correlation method for estimation of time delay. IEEE Transactions on Acoustics, Speech, and Signal Processing, **ASSP-24**:320—327.

Kraft TW, Schneeweis DM, Schnapf JL (1993) Visual transduction in human rod photoreceptors. J Physiol. **464**:747-65.

Levin N, Orlov T, Dotan S, Zohary E (2006) Normal and abnormal fMRI activation patterns in the visual cortex after recovery from optic neuritis. Neuroimage. **33**:1161-8.

Likova LT, Tyler CW (2007) Instantaneous stimulus paradigm: Cortical network and dynamics of figure-ground organization. Human Vision and Electronic Imaging XII **6492**: 248-255.

Logothetis NK (2002) The neural basis of the BOLD fMRI signal. Phil. Trans. R. Soc. Lond. **357**:1003-1037.

Logothetis NK (2003) The underpinnings of the BOLD functional magnetic resonance imaging signal. J Neurosci **23**:3963-3971.

Logothetis, NK, Wandell BA (2004) Interpreting the BOLD signal. Annual Review Physiology **66**:735-769

McIntosh AR, Gonzalez-Lima F (1994) Structural equation modeling and its application to network analysis in functional brain imaging. Hum Brain Mapp, **2**:2-22.

McIntosh AR, Grady CL, Ungerleider LG, Haxby JV, Rapoport SI, Horwitz B (1993) Network analysis of cortical visual pathways mapped with PET. J Neurosci, **14**: 655-66.

Menon RS, Luknowsky DC, Gati JS (1998) Mental chronometry using latency-resolved functional MRI. Proc Natl Acad Sci U S A, **95**:10902-7.

Metea MR, Newman EA (2006) Glial cells dilate and constrict blood vessels: a mechanism of neurovascular coupling. J Neurosci, **26**:2862-70.

Miller RG (1964) A trustworthy jackknife. Ann Math Stat, **35**:1594-1605.

Neima D, Regan D (1984) Pattern visual evoked potentials and spatial vision in retrobulbar neuritis and multiple sclerosis. Arch Neurol, **41**:198-201.

Pitman, EJG (1937) Significance tests which may be applied to samples from any population, Royal Statistical Society Supplement, **4**: 119-130 and 225-32

Regan D, Tyler CW (1971) Temporal summation and its limit for wavelength changes: an analog of Bloch's law for color vision. J Opt Soc Am, **61**:1414-21.

Regan D, Milner BA, Heron JR (1976) Delayed visual perception and delayed visual evoked potentials in the spinal form of multiple sclerosis and in retrobulbar neuritis. Brain, **99**:43-66.

Roebroeck A, Formisano E, Goebel R (2005) Mapping directed influence over the brain using Granger causality and fMRI. NeuroImage, **25**: 230-42.

Russ MO, Cleff U, Lanfermann H, Schalnus R, Enzensberger W, Kleinschmidt A (2002) Functional magnetic resonance imaging in acute unilateral optic neuritis. J Neuroimaging.,**12**:339-50.

Sato JR, Junior EA, Takahashi DY, Felix MM, Brammer MJ, Morettin PA (2006) A method to produce evolving functional connectivity maps during the course of an fMRI experiment using wavelet-based time-varying Granger causality. NeuroImage, **31**:187-96.

Schneeweis DM, Schnapf JL (1995) Photovoltage of rods and cones in the macaque retina. Science, **268**:1053-6.

Schneeweis DM, Schnapf JL (1999) The photovoltage of macaque cone photoreceptors: adaptation, noise, and kinetics. J Neurosci, **19**:1203-16.

d'Avossa G, Shulman GL, Corbetta M (2003) Identification of cerebral networks by classification of the shape of BOLD responses. J Neurophysiol, **90**:360-71.

Sekiguchi N, Williams DR, Brainard DH (1993) Efficiency in detection of isoluminant and isochromatic interference fringes. J Opt Soc Am, A **10**:2118-33.

Shank R, Aprison M (1979) Biochemical aspects of the neurotransmitter function of glutamate. In: Glutamic Acid: Advances in Biochemistry and Physiology. New York: Raven, p. 139–150.

Shmuel A, AugathM, Oeltermann A, Logothetis NK (2006) Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. Nature Neurosci, **9**, 569-577.

Sims CA (1980) Macroeconomics and reality, Econometrica. **48**, 1-48

Stefano E, Cupini LM, Rizzo P, Pierelli F, Rizzo PA (1991) Simultaneous recording of pattern electroretinogram (PERG) and visual evoked potential (VEP) in multiple sclerosis. Acta Neurol Belg, **91**:20-8.

Tolhurst DJ, Movshon JA, Dean AF (1983) The statistical reliability of signals in single neurons in cat and monkey visual cortex. Vision Res, **23**:775-85.

Tolhurst DJ, Movshon JA, Thompson ID (1981) The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast Exp Brain Res, **41**:414-9.

Uludag K, Yesilyurt B, Ugurbil K (2006) Hemodynamics and nonlinearities of BOLD response to ultrashort visual stimulation (5ms - 1s) Proc ISMRM, 5772.

Vogels R, Spileers W, Orban GA (1989) The response variability of striate cortical neurons in the behaving monkey. Exp Brain Res, **77**:432-436

Wang Y, Floor E (1994) Dynamic storage of glutamate in rat brain synaptic vesicles. Neurosci Lett, **180**:175–178.

Werring DJ, Bullmore ET, Toosy AT, Miller DH, Barker GJ, MacManus DG, Brammer MJ, Giampietro VP, Brusa A, Brex PA, Moseley IF, Plant GT, McDonald WI, Thompson AJ (2000) Recovery from optic neuritis is associated with a change in the distribution of cerebral response to visual stimulation: a functional magnetic resonance imaging study. J Neurol Neurosurg Psychiatr,. **68**:441-9.