Nonlinear Event-Related Responses in fMRI

Karl J. Friston, Oliver Josephs, Geraint Rees, Robert Turner

This paper presents an approach to characterizing evoked hemodynamic responses in fMRI based on nonlinear system identification, in particular the use of Volterra series. The approach employed enables one to estimate *Volterra kernels* that describe the relationship between stimulus presentation and the hemodynamic responses that ensue. Volterra series are essentially high-order extensions of linear convolution or "smoothing." These kernels, therefore, represent a nonlinear characterization of the hemodynamic response function that can model the responses to stimuli in different contexts (in this work, different rates of word presentation) and interactions among stimuli. The nonlinear components of the responses were shown to be statistically significant, and the kernel estimates were validated using an independent event-related fMRI experiment. One important manifestation of these nonlinear effects is a modulation of stimulus-specific responses by preceding stimuli that are proximate in time. This means that responses at high-stimulus presentation rates saturate and, in some instances, show an inverted U behavior. This behavior appears to be specific to BOLD effects (as distinct from evoked changes in cerebral blood flow) and may represent a hemodynamic "refractoriness." The aim of this paper is to describe the theory and techniques upon which these conclusions were based and to discuss the implications for experimental design and analysis.

Key words: nonlinear system identification; functional neuroimaging; fMRI; hemodynamic response function; Volterra series.

INTRODUCTION

This paper is about evoked hemodynamic responses in functional magnetic resonance imaging (fMRI) and how the measured blood oxygen level dependent (BOLD) effects can be related to underlying neuronal activity. In particular we investigate the nonlinear nature of this response, the significance of the nonlinear components, and how they affect the design and interpretation of fMRI experiments.

In Friston *et al.* (1), we presented a model of observed hemodynamic responses, in fMRI time series, that obtain when the underlying neuronal activity (inferred on the basis of changing task conditions) is convolved or smoothed with a *hemodynamic response function.* This model was subsequently elaborated in the context of the general linear model (2–4). The general linear model is variously employed in linear system identification, from a signal-processing perspective, or multiple linear regression or AnCova in statistics. These approaches to modeling and characterizing fMRI time series are predicated on the assumption that the relationship between evoked neuronal activity and the ensuing hemodynamic response can be approximated by a linear convolution using a fixed and time-invariant hemodynamic response function (1, 5). This assumption has been evaluated by comparing estimates of the hemodynamic response function using different stimuli (6) and linear system identification. In this paper, we present a nonlinear characterization of the hemodynamic response function using nonlinear system identification and explicitly assess both the significance and behavior of the nonlinear components (where they exist).

We have employed a parametric experimental design using evoked responses to words presented aurally at varying frequencies. This variation allowed us to examine the "stability" of the hemodynamic response to single words and the interactions among stimuli when presented close together. These interactions represent nonlinear effects that we were able to estimate and make statistical inferences about. The importance of this work relates to understanding the nonlinear relationship between evoked responses and sensory or behavioral parameters (such as presentation rate) and the implied constraints imposed upon experimental design and analysis. For example, we were able to resolve the apparent discrepancy between linear increases in blood flow in response to increasing word presentation rates (7) and the nonlinear dependency of the BOLD signal (8). From a data analysis perspective, the framework described in this paper can be seen as a generalization of linear approaches that characterize hemodynamic responses, evoked by single events, in terms of basis functions of peri-stimulus time (9).

This paper is divided into three sections. The first section describes the theoretical background to analyzing nonlinear or dynamic systems using Volterra series as a general model relating changes in neuronal activity (or stimuli) to hemodynamic responses. By using a second-order expansion, we were able to reformulate the problem in terms of the general linear model and therein provide for both parameter estimation and statistical inference about the effects observed. The second section uses the results of the first section to analyze two fMRI experiments, comprising two single-subject aural-stimulation paradigms. In the first, *epoch-related* experiment blocks, or epochs, of words were presented at different rates. On the basis of this experiment, we were able to estimate a high-order or nonlinear hemodynamic response function and use it to predict the response that would have been evoked by a single word. In the second, *event-related* experiment, words were presented in isolation. This allowed us to validate the estimated responses to single words from the first study in terms of

MRM 39:41–52 (1998)

From the Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London, United Kingdom.

Address correspondence to: Karl J. Friston, The Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London, UK WC1N 3BG.

Received January 21, 1997; revised June 17, 1997; accepted June 17, 1997.

This work was supported by the Wellcome Trust.

0740-3194/98 $3.00

empirically determined event-related responses from the second. The third section uses the nonlinear model of evoked responses of the previous sections to look in detail at the nonlinear interactions. By performing "virtual" experiments on the model, we show that these interactions can be thought of in terms of a hemodynamic "refractoriness" in which a prior stimulus modulates the response to a subsequent stimulus, if it occurs within a second or so. This modulation represents an interaction, over time, between the responses to successive stimuli and results in reduced responsiveness at high-stimulus frequencies.

THEORETICAL BACKGROUND

Nonlinear System Identification

Neuronal and neurophysiological dynamics are inherently nonlinear and lend themselves to modeling by nonlinear dynamic systems. However, due to the complexity of biological systems, it is difficult to find analytic equations that describe them adequately [although see Vazquez and Noll (10) for a compelling example]. An alternative is to take a very general model and obtain the specific parameters that enable the model to describe the system in question (11). A common example of this functional approach to system identification is the use of Volterra series. The Volterra series is an extension of the Taylor series representation to cover dynamic systems and has the general form

$$\begin{aligned} y(t) = {} & h^0 \\ & + \int_{-\infty}^{\infty} h^1(\tau_1) \cdot u(t-\tau_1) d\tau_1 \\ & + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h^2(\tau_1,\tau_2) \cdot u(t-\tau_1) \cdot u(t-\tau_2) d\tau_1 d\tau_2 \\ & + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h^n(\tau_1, \ldots \tau_n) \\ & \quad \cdot u(t-\tau_1) \ldots u(t-\tau_2) d\tau_1 \ldots d\tau_n \\ & + \ldots \text{ and so on} \end{aligned}$$

$y(t)$ is the output, in this case the hemodynamic response or fMRI signal, and $u(t)$ the input, in this case neuronal activity as indexed by the stimulus rate employed in our experiments. $h^n(\tau_1, \ldots \tau_n)$ is the nth order Volterra kernel. It can be shown that these series can represent any analytic time-invariant system (11). The Volterra series has been described as a "power series with memory" [see Chapters 2 and 3 of Bendat (12) for a fuller discussion]. The problem of characterizing the relationship between the stimulus function or neuronal activity $u(t)$ and the hemodynamic response $y(t)$ reduces to estimating the kernel coefficients h^n. For long time series, with relatively noiseless data, a number of approaches could be tried. Wray and Green (11) describe a technique using time-delay neuronal networks. In our experience, the nature of fMRI data does not permit the use of such techniques, so we have adopted a standard least squares approach. This has the advantage of providing for statistical inference using the general linear model (see below). To do this, one must first linearize the problem. Consider the second-order approximation to the above expansion with finite "memory" T:

$$\begin{aligned} y(t) \approx {} & h^0 \\ & + \int_0^T h^1(\tau_1) \cdot u(t-\tau_1) d\tau_1 \\ & + \int_0^T \int_0^T h^2(\tau_1,\tau_2) \cdot u(t-\tau_1) \cdot u(t-\tau_2) \cdot d\tau_1 d\tau_2 \end{aligned} \quad [1]$$

Note that the integrals start at zero. This reflects the fact that our system is "causal" in the sense that neuronal changes precede hemodynamic responses. In this formulation, the first-order coefficients $h^1(t)$ correspond to the [linear] hemodynamic response function as described in Friston *et al.* (1). The new terms depend on the second-order coefficients h^2 and are the primary focus of this paper.

The second step in making the estimates of h^0, h^1, and h^2 more tractable, for noisy data like fMRI, is to expand the kernels in terms of a small number P of temporal basis functions $b_i(\tau_1)$. This allows us to estimate the coefficients of this expansion using standard least squares: 97).

$$\begin{aligned} \text{let} \quad h^0 &= g^0 \\ h^1(\tau_1) &= \sum_{i=1}^{P} g_i^1 b_i(\tau_1) \\ h^2(\tau_1,\tau_2) &= \sum_{i=1}^{P}\sum_{j=1}^{P} g_{ij}^2 b_i(\tau_1) \cdot b_j(\tau_2) \end{aligned} \quad [2]$$

Now define a new set of response variables $x_i(t)$ that represent the original time series $u(t)$ convolved with the ith basis function

$$x_i(t) = \int b_i(\tau_1) \cdot u(t-\tau_1) d\tau_1$$

Substituting this expression into Eq. [1] and including an explicit error term $e(t)$ gives

$$y(t) = g^0 + \sum_{i=1}^{P} g_i^1 x_i(t) + \sum_{i=1}^{P}\sum_{j=1}^{P} g_{ij}^2 x_i(t) \cdot x_j(t) + e(t) \quad [3]$$

This is simply a general linear model with response variable $y(t)$, the observed time series, and explanatory variables 1, $x_i(t)$, and $x_i(t).x_j(t)$ at the [discrete] times at which they are observed. These explanatory variables (convolved time series of neuronal activity or stimulus presentation rate) constitute the columns of the *design*

matrix. The unknown parameters are g^0, g^1, and g^2 from which the kernel coefficients h^0, h^1, and h^2 are derived, using Eq. [2]. Having reformulated the problem in this way, we can now use standard analysis procedures developed for serially correlated fMRI time series that employ the general linear model (2, 4). These procedures provide parameter estimates (i.e., estimates of the basis function coefficients and, implicitly, the kernels themselves) and statistical parametric maps (SPMs) testing the significance of a hemodynamic response at each and every voxel. In this paper, we will use SPMs of the F statistic (SPM{F}) that test the joint contribution of effects considered of interest (the remaining effects, or columns of the design matrix, are called confounds). Below we will present SPM{F}s testing for the significance of the first- and second-order coefficients h^1 and h^2 and SPM{F}s that test for the nonlinear effects h^2 alone by treating the first-order effects h^1 as confounds. These analyses involve using design matrices with and without the effects of interest and assessing the reduction in error with the F statistic.

In this work, we used only three basis functions, i.e., $P = 3$ (Fig. 1). These were gamma density functions peaking during the early, intermediate, and late components of the anticipated hemodynamic response. The choice of these functions was motivated by prior knowledge about the form of the [linear] hemodynamic response function. This form is usually well approximated by a linear combination of two or more gamma density functions. A special case of this is the Poisson form adopted in Friston *et al.* (1) that corresponds to a single gamma density with equal mean and variance. Clearly the choice of basis functions is dictated by the nature of the data and the amount of temporal detail that one wants to model. In some instances (e.g., multislice acquisition), there are differences in the times that one voxel time series is acquired in relation to another. To accommodate these slight shifts in time, we often supplement the basis functions with their temporal derivatives (Fig. 1). The role of these derivatives can be seen intuitively by noting that adding (or subtracting) the temporal derivative shifts the basis function backwards (or forwards) in time. In this paper, derivatives were only used in the analysis of the event-related study where temporal effects were more acute.

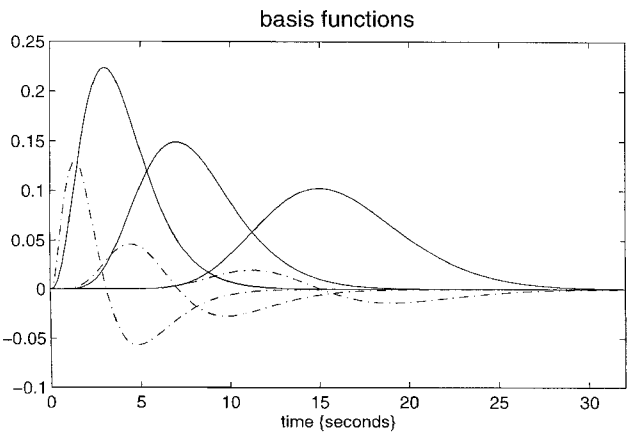

FIG. 1. Basis functions $b_i(t)$ (solid lines) and their derivatives (broken lines) used in the expansion of the Volterra kernels $h^1(\tau_1)$ and $h^2(\tau_1, \tau_2)$. These are gamma density functions with mean and variance 2^i ($i = 2$, 3, and 4). These gamma density functions can be thought of as a set of time-scaled Poisson functions because their mean and variance are equal.

THE fMRI EXPERIMENTS

Experimental Design and Data Acquisition

In this section, we apply the theory presented above to fMRI time series obtained from a single normal male subject during passive listening to words presented alone or continuously at different rates. The data were acquired at 2 Tesla using a Magnetom VISION (Siemens, Erlangen) whole body MRI system, equipped with a head volume coil. Contiguous multislice T_2*-weighted fMRI images were obtained with a gradient echo-planar sequence using an axial slice orientation (TE = 40 ms, TR = 1.7 s, 64 × 64 × 16 voxels [19.2 × 19.2 × 4.8 cm]). After discarding initial scans (to allow for magnetic saturation effects), each time series comprised 1200 (first study) and 1000 (second study) volume images with 3-mm isotropic voxels. In the first, epoch- or rate-related experiment, the subject listened to monosyllabic or bisyllabic concrete nouns (i.e., dog, radio, mountain, gate) presented at five different rates (10, 15, 30, 60, and 90 words/min) for epochs of 34 s (20 scans), intercalated with periods of rest. The five presentation rates were successively repeated according to a Latin Square design. In the second, event-related study, the subject listened to (nonrepeating) nouns presented once every 34 s.

Data Preprocessing

The data were analyzed with SPM96 (Wellcome Department of Cognitive Neurology, http://www.fil.ion.ucl.ac.uk/spm). The time series were realigned, corrected for movement-related effects, and spatially normalized into the standard space of Talairach and Tournoux (13) using the subject's coregistered structural T_1 scan (14, 15). The data were spatially smoothed with a 5-mm isotropic Gaussian kernel and temporally smoothed with a $\sqrt{8}$-s Gaussian kernel. Because we also smoothed the design matrix, the temporal smoothing does not affect the kernel or response function estimates (2).

Epoch-Related Responses

The data were analyzed using a design matrix that included the explanatory variables (convolved time series) in Eq. [3]. The basis functions employed in this analysis were a series of gamma density functions as shown in Fig. 1 (solid lines). The stimulus function $u(t)$, the supposed neuronal activity, was simply the word presentation rate at which the scan was acquired. We also used more comprehensive forms for $u(t)$ that involved modeling each word individually, but the results were very similar to the simpler analysis presented here. The resulting SPM{F}, reflecting the significance of an evoked response (or more formally, testing the null hypothesis that all h^1 and h^2 were jointly zero), is shown in Fig. 2

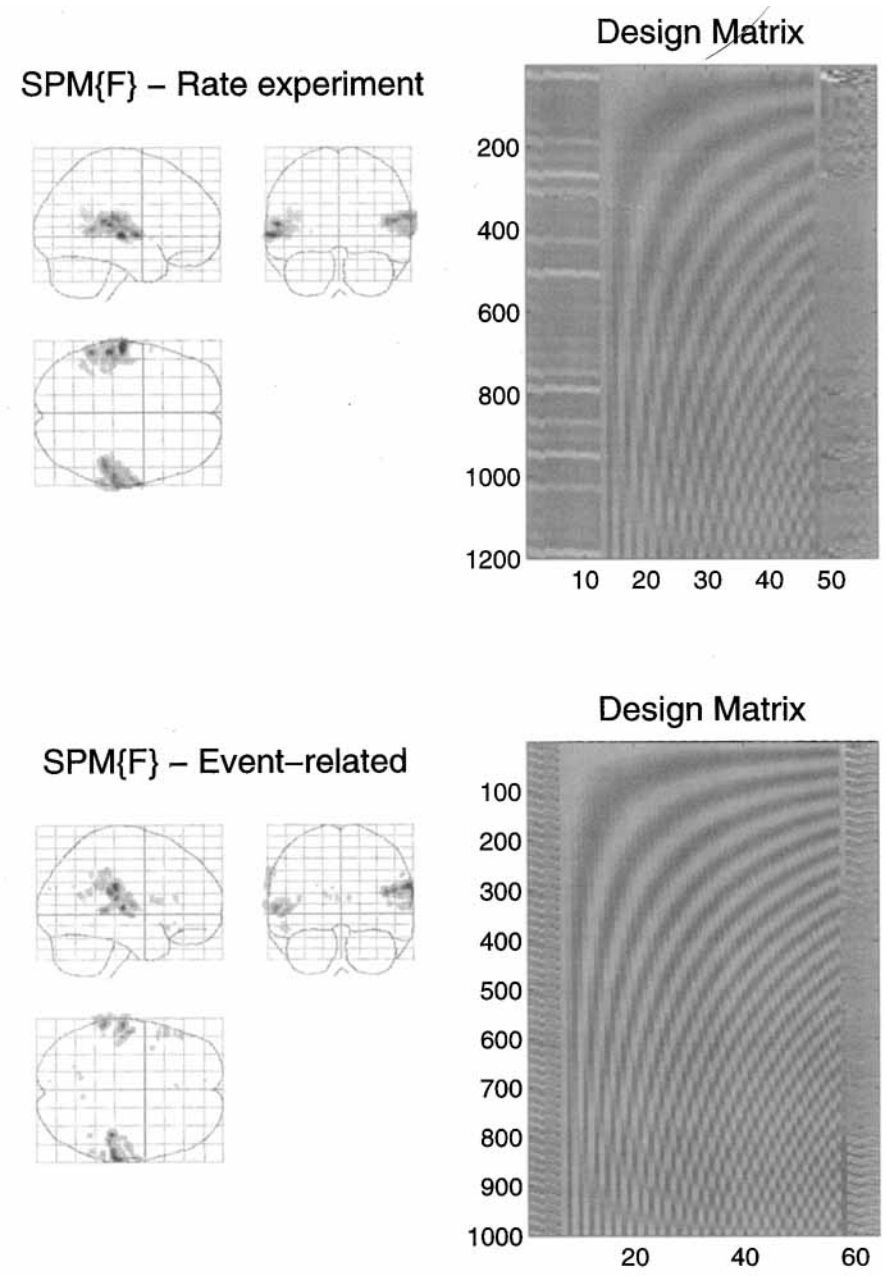

FIG. 2. Top left: SPM{F} testing for the significance of the first- and second-order kernel coefficients (h^1 and h^2) in the first (rate) experiment. This is a maximum intensity projection of a statistical process of the F ratio, following a multiple regression analysis at each voxel. The format is standard and provides three orthogonal projections in the standard space conforming to that described in Talairach and Tournoux (13). The grey scale is arbitrary, and the SPM{F} has been thresholded at 32 ($P < 0.001$ corrected). Top right: The design matrix used in the analysis. The design matrix comprises the explanatory variables in the general linear model. It has one row for each of the 1200 scans and one column for each explanatory variable or effect modeled. The left-hand columns contain the explanatory variables of interest $x_i(t)$ and $x_i(t).x_j(t)$, where $x_i(t)$ is word presentation rate $u(t)$ convolved with the basis functions $b_i(t)$ in Fig. 1. The remaining columns contain covariates or effects of no interest designated as confounds. These include (left to right) a constant term (h^0), periodic (discrete cosine set) functions of time, to remove low-frequency artifacts and drifts, global or whole brain activity $G(t)$, and interactions between global effects and those of interest $G(t).x_i(t)$ and $G(t).x_i(t).x_j(t)$. The latter confounds remove effects that have no regional specificity. Lower left: SPM{F} as above but for the second event-related experiment. This SPM{F} has been thresholded at $F = 16$ ($P < 0.001$ corrected). Note the similarity between the two SPM{F}s, despite the fact that they derive from different experimental designs and completely independent data. Lower right: The design matrix employed. The effects of interest are the first-order terms $x_i(t)$ derived by convolving $u(t)$ with all six functions in Fig. 1. $u(t)$ was in this instance one when a single word was presented and zero elsewhere. The confounds are as described above.

along with the design matrix used. The left-hand side of the upper design matrix comprises the explanatory variable $x_i(t)$ and $x_i(t).x_j(t)$. The remaining columns contain the constant (used to estimate g^0) and other effects designated as confounds (low-frequency artifacts, global effects, and so on). This SPM{F} has been thresholded [$F = 32, P < 0.001$ corrected for multiple comparisons (16)] and shows widespread responses in bilateral temporal regions with the most significant effects evident in the periauditory regions.

The estimated kernels h^0, h^1, and h^2 for a voxel in the left superior temporal gyrus (-56, -28, 12 mm) are shown in Fig. 3. As might be expected, the first-order kernel resembles the hemodynamic response functions identified using linear analyses such as least squares deconvolution [e.g., Fig. 6 in Ref. (1)] or linear regression [e.g., Fig. 5 in Boynton *et al.* (6)]. Of note is the protracted undershoot that lasts for about 16 s. The second-order kernel is shown below and is remarkable for the pronounced negativity on the lower left, flanked by smaller positive lobes. This negativity suggests that if neuronal activity has been high in the past few seconds, then the hemodynamic response will be suppressed. The positive lobes suggest that this suppression is ameliorated if the underlying neuronal activity is sustained, i.e., is high in the recent (4 s) and more distant past (8 s). There are two further important points to note. First, the second-order kernel is symmetrical. This will always be the case because the contribution of $u(t - \tau_1).u(t - \tau_2)$ to the response is exactly the same as $u(t - \tau_2).u(t - \tau_1)$. The second unpredicted and more intriguing observation is that the second-order kernel is very similar to the "product" of the first-order kernel times itself

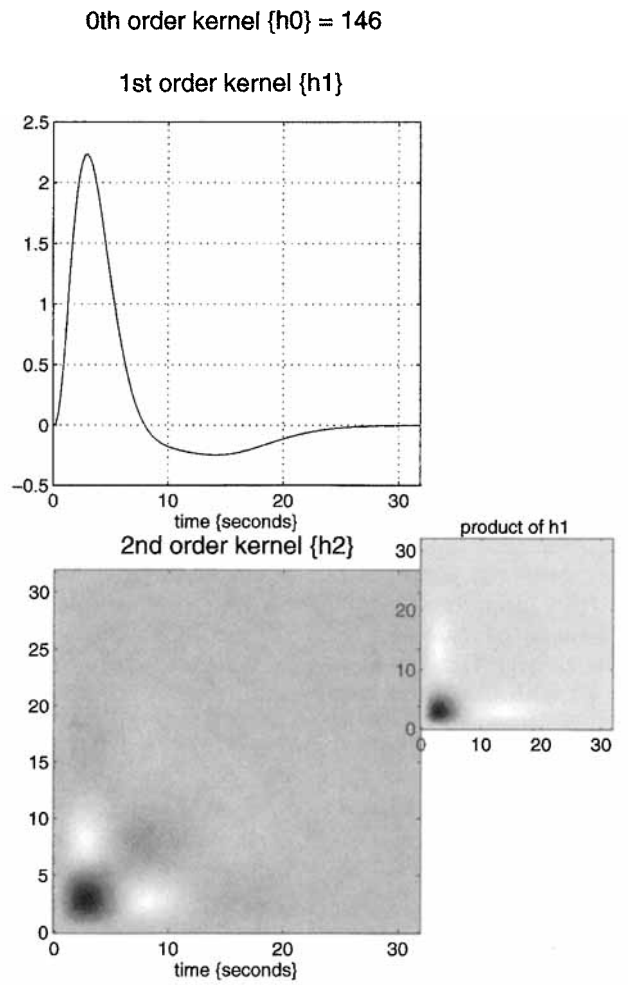

FIG. 3. The Volterra kernels h^0, h^1, and h^2 based on parameter estimates from a voxel in the left superior temporal gyrus at -56, -28, 12 mm. These kernels can be thought of as a characterization of the second-order hemodynamic response function. The first-order kernel (upper panel) represents the (first-order) component usually presented in linear analyses. The second-order kernel (lower panel) is presented in image format. The color scale is arbitrary; white is positive and black is negative. The insert on the right represents $[-h^1(\tau_1).h^1(\tau_2)]$, the second-order kernel that would be predicted by a simple model that involved convolution with h^1 and then some nonlinear scalar function.

(insert in Fig. 3). In other words, $h^2(\tau_1,\tau_2)$ is roughly proportional to $h^1(\tau_1).h^1(\tau_2)$. We return to the implications of this in a subsequent section.

The basic form of the second-order kernel estimates was very similar for all the voxels in the periauditory region. The nature of these nonlinear effects will be demonstrated more intuitively in the final section of this paper. Using these kernel estimates, we can estimate responses to any temporal pattern of words presented by using Eq. [1] and any suitable function $u(t)$. In the first instance, we present the estimated responses to the stimuli actually used: The predicted and observed responses for the first 17 min of the time series are shown in Fig. 4 (upper panel). Predicted responses to 34-s epochs at two presentation rates (30 and 60 words/min) and the adjusted responses observed are shown in the lower panel of Fig. 4 (adjusted data is simply the original data after the confounding effects have been removed). The agreement is evident.

Event-Related Responses

By specifying a stimulus function $u(t)$ that models the occurrence of a single word, we can use Eq. [1] and the kernel estimates in Fig. 3 to simulate the hemodynamic response of this brain region to single word. This simulated event-related response is shown in Fig. 5 (upper panel). One observes a peak at about 4 s followed by a protracted undershoot lasting for about 16 s. This response is "simulated" using a model whose parameters were determined without ever presenting single words in isolation (i.e., the Volterra series model based on the rate experiment). A validation of the model can be effected in terms of the empirically determined event-related response to actual single words using the second experiment.

The same analysis described above was applied to the event-related, single word experiment. In this instance, by virtue of the fact that the words were presented very sparsely, there is no opportunity for the responses to

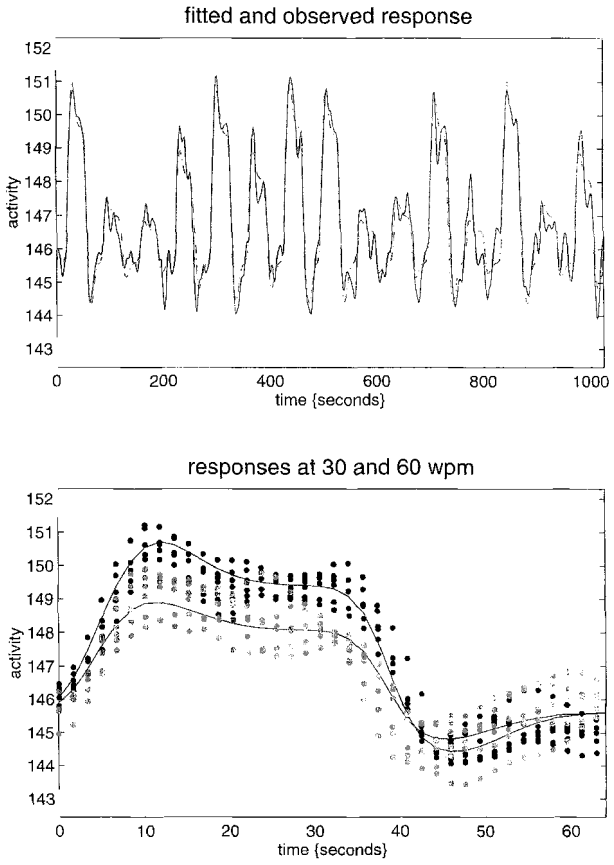

FIG. 4. Top panel: Fitted or predicted (broken line) and observed (solid line) responses at the same voxel as in Fig. 3, over the first 1024 s. The observed responses here are adjusted such that those effects that can be modeled by the confounds have been removed. Lower panel: Predicted and observed responses for epochs of 30 (light grey) and 60 (dark grey) words/min (wpm) superimposed upon each other.

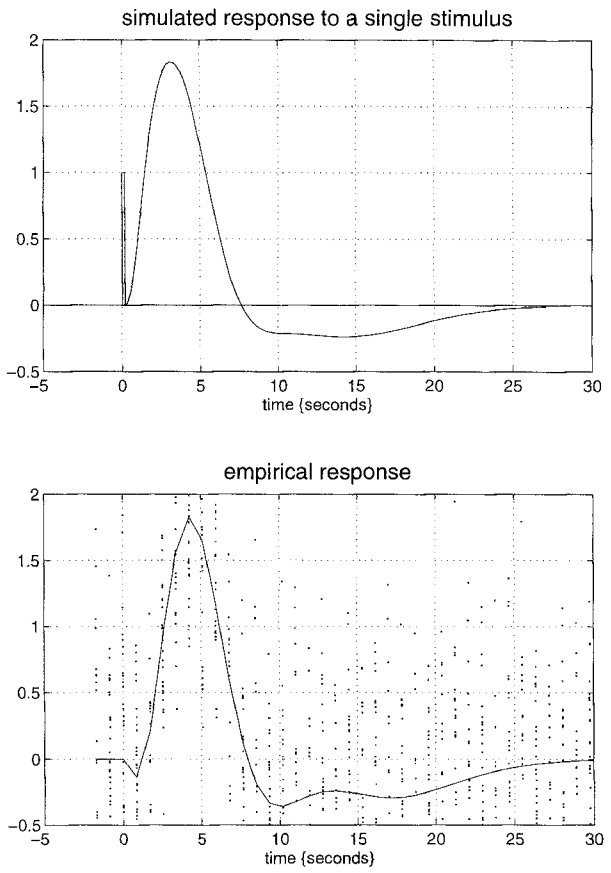

FIG. 5. Top panel: Hemodynamic response to a single word (bar at 0 s) modeled using the Volterra kernel estimates of Fig. 3. Lower panel: The empirical event-related response in the same region based on the second experiment. The solid line is the fitted response using the first-order kernel estimates and the dots represent the adjusted responses.

successive words to interact, and therefore a first-order model is sufficient to describe the response. In this case $u(t)$ was effectively a series of delta functions modeling the occurrence of each word. The basis functions were the gamma functions used above (solid lines in Fig. 1) and their derivatives (dotted lines in Fig. 1). The resulting SPM{F}, testing for the significance of event-related responses to single words, is shown in the lower half of Fig. 2. This should be compared to the equivalent SPM{F} from the rate experiment (upper half). The similarity between these two SPM{F}s is remarkable given that they were obtained using completely different data and experimental designs (epoch- or rate-related and event-related). The fitted response, based on the estimate of h^1, from the same region as above, is shown in Fig. 5 (lower panel) with the adjusted data. The striking similarity between the empirically observed event-related response and that predicted on the basis of the Volterra kernels h^0, h^1, and h^2 obtaining from the rate experiment (upper panel) can be considered a validation of the estimation procedure and the underlying model. Interestingly the empirical response includes a slight initial "dip." We will return to this below.

NONLINEAR ASPECTS OF EVOKED RESPONSES

Are They Significant?

To assess the significance of the nonlinear response components (due to h^2), over and above the first-order components, we repeated the analysis of the rate experiment, treating the first-order effects (i.e., the contributions determined by h^1) as confounds. The resulting SPM{F} is shown in Fig. 6 and implicates both periauditory regions and the left posterior superior temporal region (Wernicke's area), suggesting that nonlinear effects are not only prevalent but very significant ($P < 0.001$ corrected) in this experimental design.

The Form of the Kernel Estimates and Implications for the "Structure" of Nonlinear Effects

As noted above and brought to our attention by one of our reviewers, the elements of the second-order kernel are roughly proportional to the product of the corresponding elements of the first-order kernel, i.e., $h^2(\tau_1,\tau_2) \propto h^1(\tau_1).h^1(\tau_2)$. This can be seen by comparing the estimate of h^2 with the insert corresponding to $-h^1(\tau_1).h^1(\tau_2)$ in Fig. 3. There are subtle differences in that the off-diagonal positive lobes in h^2 peak around 8 s, whereas they peak at 16 s in the insert (at these times the values of h^2 are slightly negative). However, the overall form is very similar and this suggests a simple form for the underlying nonlinear model of evoked responses:

$$y(t) \approx f\left(\int_0^T h^1(\tau_1) \cdot u(t-\tau_1)d\tau_1\right)$$

where $f(.)$ is a nonlinear scalar function. Expansion of $f(.)$ in a McLaurin series gives

$$\begin{aligned} y(t) \approx\ & f(0) \\ & + f'(0)\int_0^T h^1(\tau_1) \cdot u(t-\tau_1)d\tau_1 \\ & + \frac{f''(0)}{2}\int_0^T\int_0^T h^1(\tau_1) \cdot h^1(\tau_2) \cdot u(t-\tau_1) \\ & \cdot u(t-\tau_2) \cdot d\tau_1 d\tau_2 \end{aligned} \qquad [4]$$

Eq. [4] demonstrates the formal similarity with Eq. [1] where $h^2(\tau_1,\tau_2)$ has been replaced by $h^1(\tau_1).h^1(\tau_2)$. This simpler model is equivalent to convolving the stimulus function with a first-order kernel (i.e., a linear "latent" hemodynamic response function) and then taking some nonlinear (e.g., second-order polynomial) function of the result. The distinction between the general form implied by the Volterra series and this simpler form is depicted in Fig. 7. It is pleasing to note that this simple form was adopted by Vazquez and Noll (10) in their nonlinear characterization of evoked visual responses. These authors assumed a Gaussian form for the kernel, but still

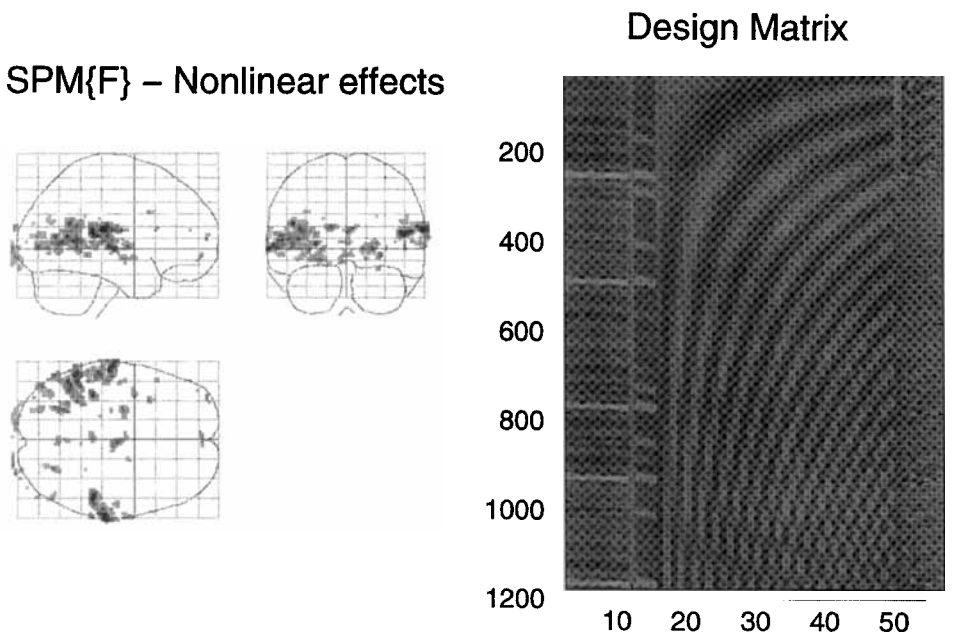

FIG. 6. Left: SPM{F} testing for the significance of the second-order kernel coefficients (h^2). In this instance, the explanatory variables pertaining to the first-order kernel have been moved over to the confound partition of the design matrix (shown on the right). The SPM{F} has been thresholded at $F = 16$ ($P < 0.001$ corrected). Note that this SPM{F} has been thresholded at half the value employed for the equivalent SPM{F} testing for the first- *and* second-order effects in Fig. 2. This is because the values in this SPM{F} were generally smaller, although still extremely significant.

our results speak to the appropriateness of the general form for their model.

From the perspective of the current analysis, the parameter estimates suggest a specific form for the nonlinear relationship between input and measured output in fMRI, despite the fact that no constraints (other than the basis functions) were placed on the model. These sorts of insights may help to identify the biophysical level at which nonlinearities are expressed in fMRI. For example, if the nonlinear effects manifest after a temporal convolution of neuronal activity to give a hemodynamic response, then it may be that the measured fMRI signal is simply a saturating (nonlinear) function of hemodynamic changes. We return to this in the discussion. Simple nonlinear forms are also important from the point of view of system identification using optimization techniques, because there are fewer parameters to estimate (and their relationship to the system in question is often more apparent). However, it should be noted that the simplification implicit in Eq. [4] does not help in the context of framework adopted here. This is because to make Eq. [4] linear in the parameters, one comes back to Eq. [3].

Interactions Between Stimuli

In this section, we examine how the nonlinear effects identified in the previous sections come to shape the responses to different stimuli. We have chosen to do this in terms of the effect that a preceding stimulus has on the response to a current stimulus. This captures the essence of nonlinear responses, in the sense that interesting nonlinearities (above and beyond a simple nonlinear mapping from neuronal input to hemodynamic response) involve interactions over time. The kernel estimates of the previous section can be used to specify a model that should reproduce the hemodynamic response to any arbitrary stimulus. This means that we can look in detail at responses to different sequences of events that would otherwise require a whole series of experiments. Of course, these simulated responses are only predictions and suggest some interesting experiments for empirical verification. Here, however, we use these predictions to convey, in a heuristic way, the implications of the second-order effects.

Consider the response to a pair of stimuli, separated by a second, in relation to the responses to each presented in isolation. Figure 8 shows these responses to the stimuli (upper panel: together–solid line, and separately–broken lines). To assess the impact of the first stimulus on the response to the second, we can subtract the response to the first stimulus from the response to both. This gives the response to the second stimulus in the context of the first (solid line in the lower panel of Fig. 8). It can be seen that this response is attenuated markedly, with an augmented undershoot, in relation to the response obtained when the stimulus is presented in isolation (broken line). In short, the response to a stimulus is compromised or modulated by preceding stimuli to give a nonlinear "refractoriness" that depends on the interstimulus interval or rate. This effect is only one aspect of the nonlinear interactions embodied in the characterization, but it is an important one. The consequence of this effect is to progressively moderate the response to each word with increasing rates of presentation. Figure 9 (upper panel) shows the responses, to epochs of words presented at different rates, predicted by the model. The responses here are simply the integral under the evoked response curve during word presentation. The empirical response (dots) are included for comparison. As expected [and consistent with the results of Binder *et al.* (8)], the response function deviates from a linear relationship at higher event frequencies. It is interesting that for the voxel we have been using, the nonlinear model predicts that the integrated response would fall off at very high rates or frequencies. This is not an artifact. In some brain regions, this effect was observed empirically. The lower panel depicts the modeled and empirical integrated responses for a voxel more anteriorly in the superior temporal gyrus that shows an inverted U dependency on presentation rate, peaking at 1/s. The neurophysiological mechanisms that may contribute to this effect are discussed below.

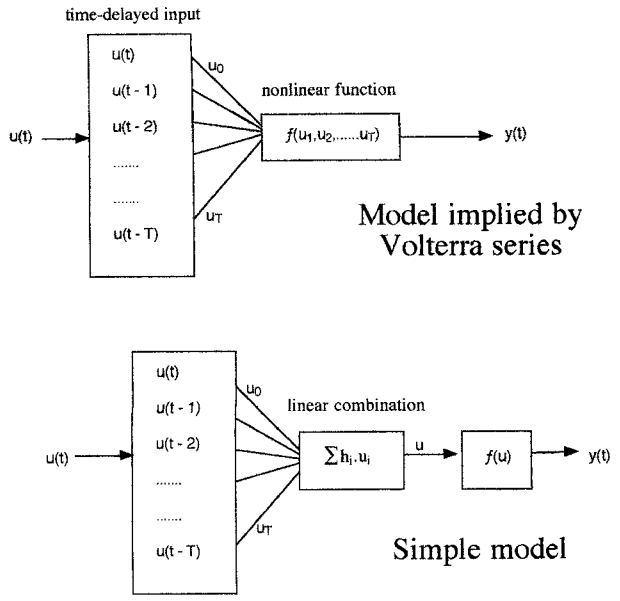

FIG. 7. Schematic depicting the difference between the finite-memory power series implied by the Volterra series model and the simpler model suggested by the parameter estimates. $u(t)$ represents the input (in this case the stimulus function) and $f(u_1,u_2,\ldots u_n)$ some nonlinear function with n arguments. $y(t)$ is the output (in this instance the hemodynamic response).

Variation in Responses Over the Brain

So far we have focussed on evoked responses at one point in the brain. In general, the kernel estimates for other brain regions were very similar in form. To characterize the variability in responses over different regions, we simulated the response to a single word at all voxels surviving an uncorrected F value corresponding to $P = 10^{-8}$ and then performed a principal component analysis of the resulting responses. The eigenvalue spectrum (Fig. 10, upper panel) suggested that there were three main response forms or components that could largely account for the variation in responses. These first three principal components are shown in the lower panel. The first corresponds to the canonical hemodynamic response with a very quick onset and early peak at 3 s. The second can be interpreted in terms of a late response, peaking at about 8 s. Interestingly, the second component shows a pronounced early "dip" and is almost exactly the same as the responses shown in Le and Hu (17). This early signal decrease has been attributed to increased oxygen extraction (18) before compensatory increases in blood flow become established (19). The third component is a late component that covers most of the "undershoot." The similarity between these principal components and the basis functions employed in the model is expected given that (i) the form of the response is constrained by the basis functions, and (ii) the basis functions were chosen to "cover" likely variations in the response profile. To indicate the spatial organization of the response variation, Fig. 11 presents the positive and negative first principal components scores as maximum intensity projections. As might be expected, the high positive scores are most evident in the auditory and periauditory regions (upper panel). The negative scores are most pronounced in posterior temporal, parietal, and extrastriate regions (lower panel). Negative expression of the first principal component corresponds to an evoked *deactivation* or reduction in signal. This is a real phenomenon and can be demonstrated as such by looking at the event-related responses in the posterior superior temporal region using independent data from the second study. Figure 12 shows the form of this response that can be characterized as a deactivation, peaking at about 5 s followed by a more protracted positive rebound. This unusual biphasic response is not an artifact of global or whole brain normalization (because we removed global confounds using multiple linear regression as opposed to scaling the data) and speaks to the ability of fMRI to detect decreases in BOLD signal that endure far longer than the early transients mentioned above.

DISCUSSION

We have presented a nonlinear approach to characterizing evoked hemodynamic responses in fMRI that is based

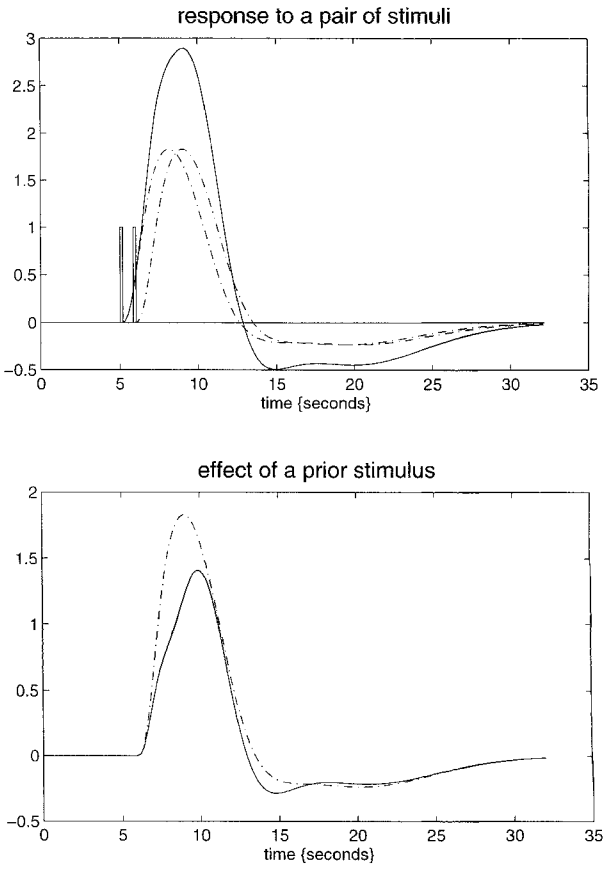

FIG. 8. Top panel: The simulated responses to a pair of words (bars) (1 s apart) presented together (solid line) and in isolation (broken line) based on the second-order hemodynamic response function in Fig. 3. Lower panel: The response to the second word when preceded by the first (broken line), obtained by subtracting the response to the first word from the response to both, and when presented alone (broken). The difference reflects the impact of the first word on the response to the second.

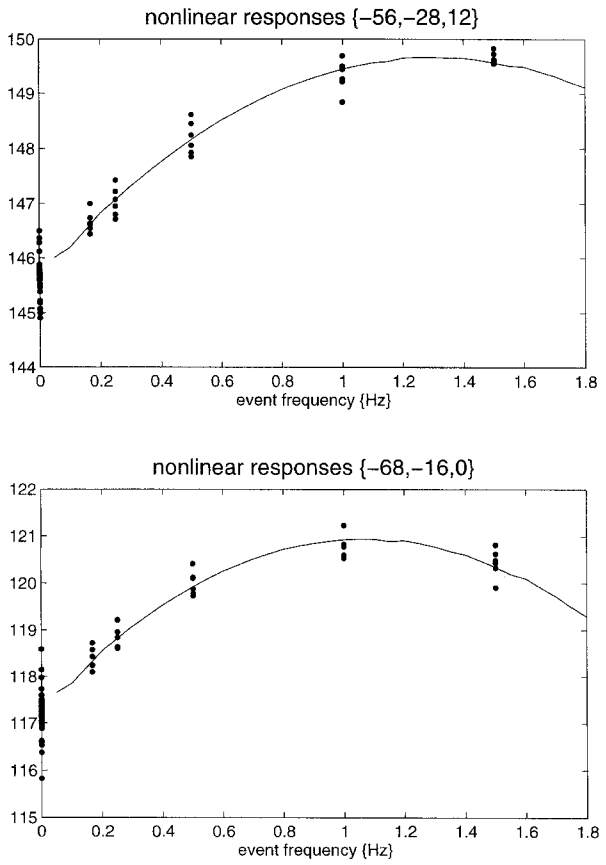

FIG. 9. Integrated (over the period of word presentation) responses during epochs of words, lasting for 34 s, presented at increasing frequencies. The line represents the simulated responses using the second-order hemodynamic response function in Fig. 3, and the dots correspond to the observed responses at the voxel in question. Lower panel: The same analysis but for a voxel more anterior in the superior temporal gyrus (-68, -16, 0 mm). Note the nonlinear and inverted U relationship between integrated response and word presentation frequency.

on nonlinear system identification, in particular the use of Volterra series. By reformulating the model, we were able to estimate the kernel coefficients that mediate between underlying neuronal activity and the observed hemodynamic response and make statistical inferences about their significance. These kernel coefficients can be thought of as high-order or nonlinear extensions of linear convolution or "smoothing" functions and, therefore, represent a nonlinear characterization of the hemodynamic response function. We have shown that the same nonlinear response function can model the responses to stimuli in different contexts (in this work, different rates of presentation) and that the nonlinear component is not only very significant but also is quantitatively important. Its effect can be thought of in terms of interactions between successive stimuli, such that the responses to an extant stimulus are modulated by the preceding stimulus. This means that responses at very high frequencies saturate and in some brain areas start to decline again. A number of techniques and observations have been presented in this paper, and we will now review and extend some of the more important issues.

Nonlinear Modeling versus Linear Modeling

What Are the Implications of this Work for Experimental Design and Analysis?

The first thing to note is that there is a fundamental distinction between positing the same nonlinear hemodynamic response function that can account for varying responses to stimuli presented at different rates, and a series of rate-dependent, linear hemodynamic response functions. To make this distinction clear, consider the analysis presented in Fig. 13. In this analysis, we have discarded the second-order terms from the design matrix and have treated each presentation rate as a different stimulus type. This simply involves separating the first-order terms into a set of columns for each rate (Fig. 12, upper right). This represents an alternative, multiple linear regression approach to the data and yields parameter estimates (up to first-order kernel coefficients h^0 and h^1) for each rate. The resulting event-related responses (shown in the lower panel of Fig. 13) are rate-specific and, as one might expect, show that responses to single words are progressively attenuated when these words are presented at high frequencies. The set of first-order hemodynamic response functions (Fig. 13) and the single

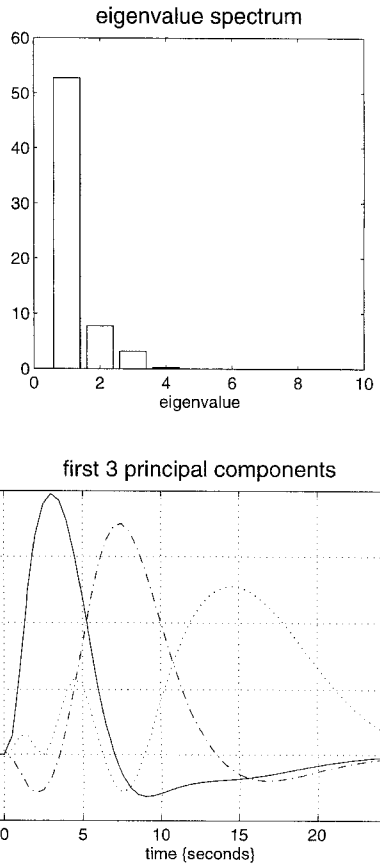

FIG. 10. Principal component analysis of simulated responses to single words in different brain regions. Top panel: Normalized eigenvalue spectrum showing only three principal components have eigenvalues greater than unity. Lower panel: The first three principal components (solid line–first, broken line–second, and dotted line–third) reflecting the underlying forms of estimated hemodynamic responses to single words.

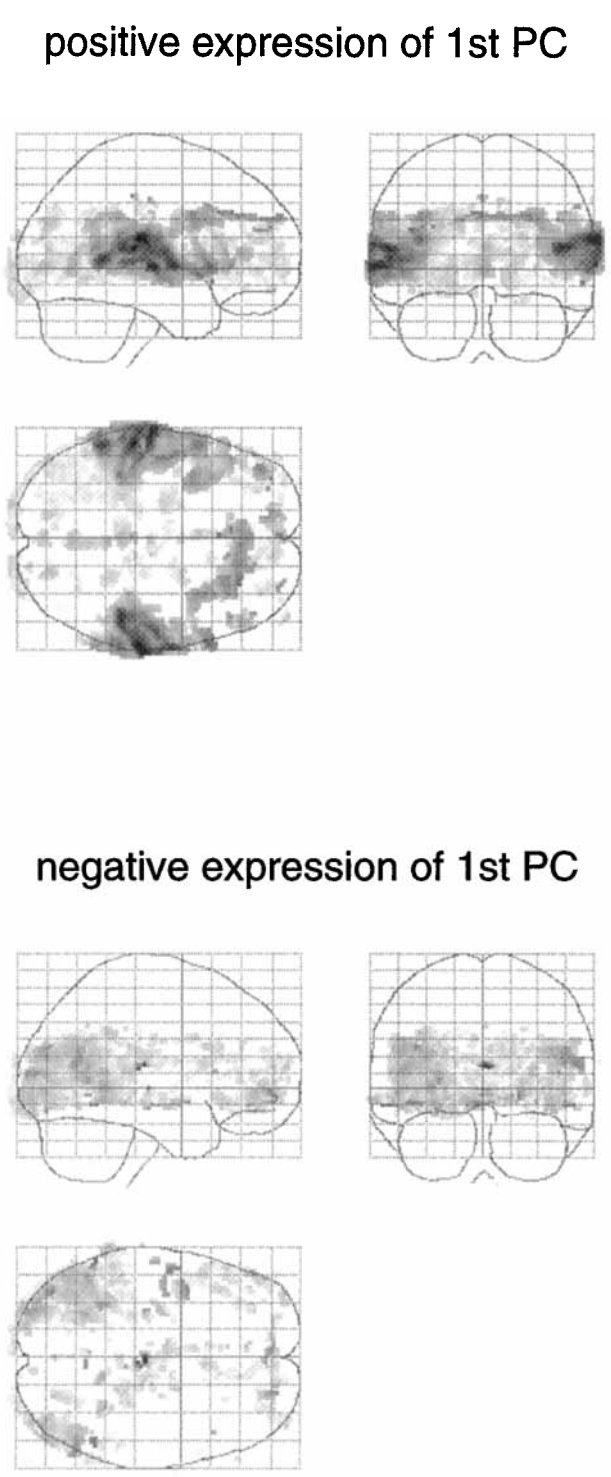

FIG. 11. Spatial distribution of component scores from Fig. 10. Top panel: Maximum intensity projection of the positive component scores showing that bitemporal, periauditory regions express the first principal component to a substantial degree. Lower panel: As above but for the negative scores associated with the first principal component of the response.

second-order hemodynamic response function (Fig. 3) are both trying to model the same thing, and yet they do so in a very distinct way. The first distinction is conceptual; in the nonlinear analysis, we are explaining the responses in terms of the same response function that accommodates interactions between stimuli. In the linear analysis we relegate these interactions to formal differences among the rate-specific response functions (and would normally try to characterize these differences *post hoc*). There is another more subtle difference between the two analyses: In the linear analysis, we have discounted second-order effects in the hope that a suitably shaped first-order response function can model all the nonlinearities inherent in the real response. While this is justifiable for epochs of a fixed rate and length, it may not be for epochs that endure over different periods of time. This is because interactions over time (e.g., hemodynamic adaptation and refractoriness) may be significant, even for a fixed stimulus frequency. Only the nonlinear analysis would model these effects appropriately. Note that for a fixed form of event, or epoch, the second-order terms [e.g., $x_i(t).x_j(t)$] can be emulated, exactly, by (more complicated) first-order terms. In other words, with no parametric variation in the form of the underlying neuronal activity evoked, a first-order model is sufficient. This is the rationale behind dropping the second-order terms in the event-related study and in the linear analysis depicted in Fig. 13.

Which Then is the Most Appropriate Analysis to Use?

This question is only posed in parametric experimental designs (20) when some experimental parameter is varied (for example, rate of stimulus presentation, response rate, duration of task, etc.). The alternatives are then to model a nonlinear response in terms of a parametrically varying input $u(t)$ or to treat each level of the parameter manipulated as a separate condition (cf. the linear analysis above). In some circumstances, only the nonlinear option may be viable, for example, if the parameter is

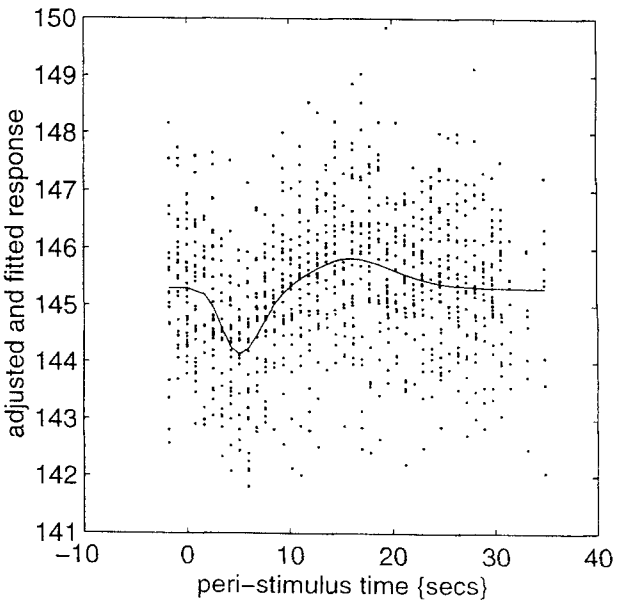

FIG. 12. Event-related deactivation: The empirical event-related response based on the zeroth and first-order kernel estimates (solid line) for a voxel in the right superior temporal region at -42, -50, 18 mm. The dots correspond to adjusted data.

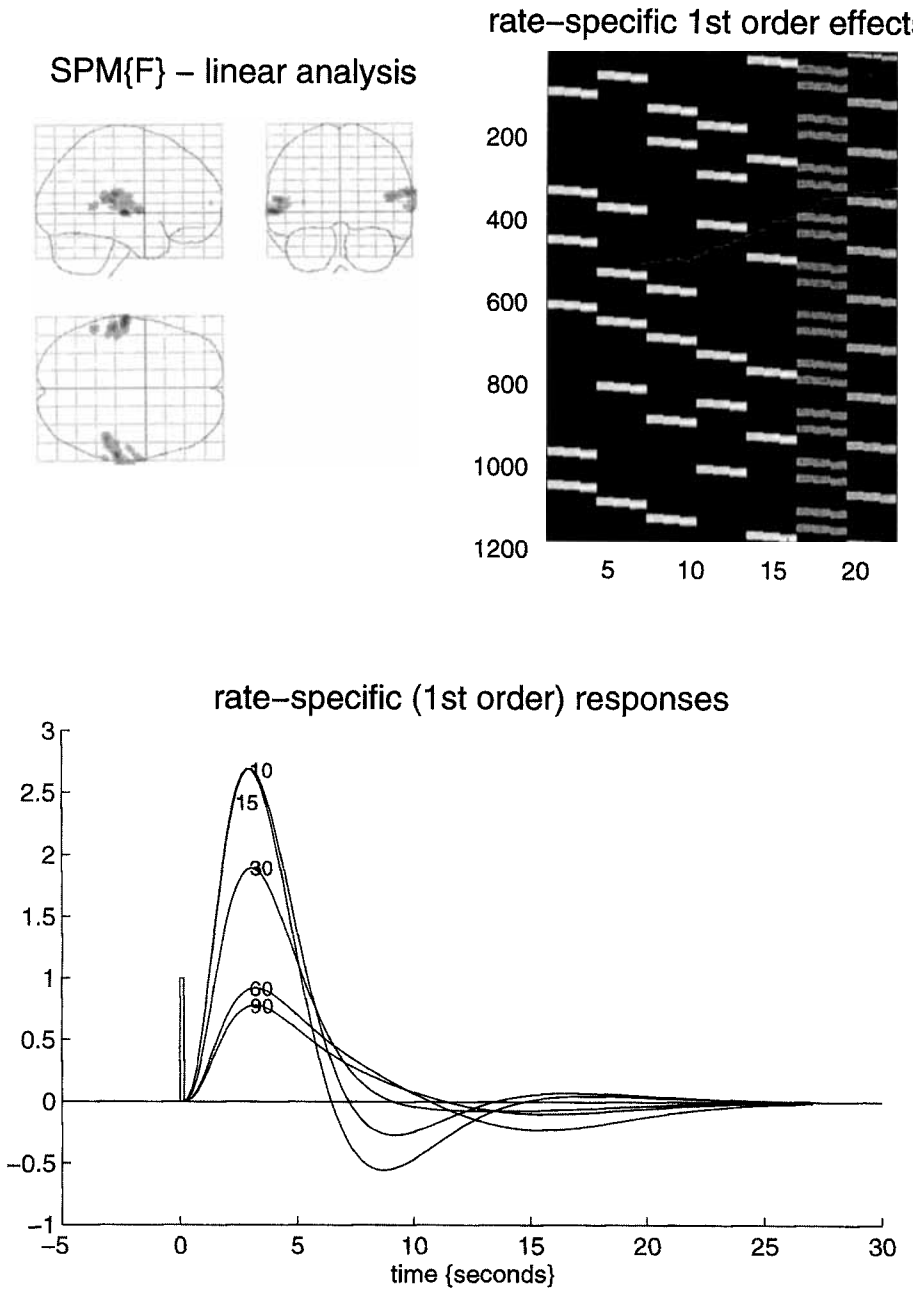

FIG. 13. A linear analysis of the rate experiment. Top left: SPM{F} testing for the significance of the first-order coefficients h^1 estimated separately (but simultaneously) for each presentation rate. This SPM{F} has been thresholded at $F = 16$ ($P < 0.001$, corrected). The similarity with the equivalent SPM{F}s in Fig. 2 is evident. Top right: Explanatory variables corresponding to $x_i(t)$ arranged to model each presentation rate independently (i.e., the design matrix partition of interest). The first five sets of columns represent the five presentation rates and the final two correspond to (short and long) periods of rest. Lower panel: Simulated hemodynamic responses to single words (bar) using the kernel estimates from the above design matrix, for the same voxel as in Fig. 3. These response-function estimates are rate-specific and show an attenuated response to single words when they are presented in the context of a high-frequency word stream. The numbers on the response functions denote the presentation rate in words/min.

changing continuously and does not conform to a series of discrete levels. In general, however, the results that obtain from the two approaches would be similar, and the question reduces to one of implementational expediency and simplicity in describing the results. Although it is possible to implement the nonlinear analysis above with existing tools (e.g., SPM96), the construction of the design matrices is complicated, and a simple linear analysis may be quite sufficient for most purposes.

In terms of experimental design, the nonlinear effects above mean that it is possible to drive the brain "too hard" with very high stimulus presentation rates. The analyses of the present study suggest that the optimum presentation rate, for words, is about 1/s.

The SPM{F}

One aspect of the techniques presented in this paper is the use of the SPM{F} to make inferences about the significance of the response, in terms of the response kernel coefficients. This is an instance of the usefulness of the SPM{F} [see Buechel *et al.* (21) for another example] and has been facilitated by recent advances in Gaussian field theory that allow one to make corrections for multiple dependent comparisons in F fields (16). The importance of the SPM{F}, as opposed to SPM{t} or t maps (and related statistical processes), is that it reflects the significance of a whole set of parameter estimates, in this instance the collection of kernel coefficients that describe the hemodynamic response (h^0, h^1, and h^2). We envisage that the SPM{F} will find an increasing role in fMRI, where the emphasis will shift from making statistical inferences about particular simple effects to quantifying their more complicated nature and form in terms of the parameter estimates. As the models of hemodynamic responses become more sophisticated, the number of parameter estimates involved will increase, and some device will be required to make an inference about these parameters *en masse*. The SPM{F} is one such device. It should be noted that the distribution of the F statistics based on temporally correlated fMRI data only approximate a true F distribution. Their use should

consequently be viewed in the same light as many other instances of inference in neuroimaging that are based on distributional approximations that are only exact in the limit of high thresholds or degrees of freedom.

Neurobiological Mechanisms

Throughout this paper, we have referred to the input $u(t)$ as underlying neuronal activity and have assumed that this is equivalent to word presentation rate. This assumption allows one to assign all the nonlinearities observed in the response to the mapping between neuronal activity and hemodynamic response. It is, of course, possible that much of the implicit refractoriness can be explained directly in neuronal terms. There is no way of distinguishing between these explanations using the present data. The reason we have assumed that neuronal activity (i.e., neuronal discharge rate) is proportional to word presentation rate is that previous observations (Ref. 7 and see below) show a proportional relationship between blood flow and word presentation rate, and blood flow is generally regarded as an index of presynaptic activity. Although neuronal adaptation, facilitation, and refractoriness certainly contribute to our results, we can make the following (very heuristic) argument. If we assume that each stimulus, or event, evokes roughly the same degree of spike activity in a neuronal population (providing that these events are at least several hundred ms apart), then spike activity and blood flow will increase in proportion to stimulus frequency. The demonstration of highly significant nonlinear components (h^2) in the hemodynamic response based on BOLD effects, therefore, suggests a nonlinear relationship between flow and oxygen extraction fraction. This nonlinear flow-dependency is fully expected (22) and represents a sufficient explanation for the observed nonlinearities. We feel compelled to address this issue because we replicated the rate experiment exactly (using the same experimental design, the same subject, and reproducing the same acoustic conditions as those experienced in the fMRI setting) using positron emission tomography. Blood flow, measured in the periauditory region, showed an almost exact linear dependency on rate. These data will be described in detail elsewhere. In short, the nonlinear components characterized by h^2 may be specific to BOLD effects.

ACKNOWLEDGMENTS

The authors thank the two anonymous reviewers for substantial help in presenting and interpreting this work.

REFERENCES

1. K. J. Friston, P. Jezzard, R. Turner, Analysis of functional MRI time series. *Hum. Brain Map.* **1,** 153–171 (1994).
2. K. J. Friston, A. P. Holmes, J.-B. Poline, P. J. Grasby, S. C. R. Williams, R. S. J. Frackowiak, R. Turner, Analysis of fMRI time-series revisited. *Neuroimage* **2,** 45–53 (1995).
3. K. J. Friston, C. D. Frith, R. Turner, R. S. J. Frackowiak, Characterizing evoked hemodynamics with fMRI. *Neuroimage* **2,** 157–165 (1995).
4. K. J. Worsley, K. J. Friston, Analysis of fMRI time-series revisited-again. *Neuroimage* **2,** 173–181 (1995).
5. P. A. Bandettini, A. Jesmanowicz, E. C. Wong, J. S. Hyde, Processing strategies for time course data sets in functional MRI of the human brain. *Magn. Reson. Med.* **30,** 161–173 (1993).
6. G. M. Boynton, S. A. Engel, G. H. Glover, D. J. Heeger, Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* **16,** 4207–4221 (1996).
7. C. J. Price, R. J. S. Wise, S. Ramsay, K. J. Friston, D. Howard, K. Patterson, R. S. J. Frackowiak, Regional response differences within the human auditory cortex when listening to words. *Neurosci. Lett.* **146,** 179–182 (1992).
8. J. R. Binder, S. M. Rao, T. A. Hammeke, J. A. Frost, P. A. Bandettini, J. S. Hyde, Effects of stimulus rate on signal response during functional magnetic resonance imaging of auditory cortex. *Cognit. Brain Res.* **2,** 31–38 (1994).
9. O. Josephs, R. Turner, K. J. Friston, Event-related fMRI. *Hum. Brain Map.* **6,** 243–248 (1997).
10. A. L. Vazquez, D. C. Noll, Non-linear temporal aspects of the BOLD response in fMRI, *in* "Proc., ISMRM, 1st Annual Meeting, 1996," p. S1765.
11. J. Wray, G. G. R. Green. Calculation of the Volterra kernels of non-linear dynamic systems using an artificial neuronal network. *Biol. Cybern.* **71,** 187–195 (1994).
12. J. S. Bendat, "Nonlinear System Analysis and Identification from Random Data," John Wiley and Sons, New York, 1990.
13. J. Talairach, P. Tournoux, "A Co-planar Stereotaxic Atlas of a Human Brain," Thieme, Stuttgart, 1988.
14. K. J. Friston, S. Williams, R. Howard, R. S. J. Frackowiak, R. Turner, Movement- related effects in fMRI time series. *Magn. Reson. Med.* **35,** 346–355 (1996).
15. K. J. Friston, J. Ashburner, C. D. Frith, J.-B. Poline, J. D. Heather, R. S. J. Frackowiak. Spatial registration and normalization of images. *Hum. Brain Map.* **2,** 165–189 (1995).
16. K. J. Worsley, Local maxima and the expected Euler characteristic of excursion sets of $\chi 2$, F and t fields. *Adv. Appl. Prob.* **26,** 13–42 (1994).
17. T. H. Le, X. Hu, Evaluation of the early response in fMRI in individual subjects using short stimulus duration, *in* "Proc., ISMRM, 1st Annual Meeting, 1996," p. S285.
18. R. S. Menon, S. Ogawa, X. Hu, 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,** 453–495 (1995).
19. D. Malonek, A. Grinvald, Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: implications for functional brain mapping. *Science* **272,** 551–554 (1995).
20. K. J. Friston, A. P. Holmes, K. J. Worsley, J.-B. Poline, C. D. Frith, R. S. J. Frackowiak, Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Map.* **2,** 189–210 (1995).
21. C. Buechel, R. S. J. Wise, C. Mummery, R. S. J. Frackowiak, K. J. Friston, Nonlinear regression in parametric activation studies. *Neuroimage* **4,** 60–66 (1996).
22. 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.,* in press (1997).