Detecting Latency Differences in Event-Related BOLD Responses: Application to Words versus Nonwords and Initial versus Repeated Face Presentations

R. N. A. Henson,*,†,1 C. J. Price,* M. D. Rugg,† R. Turner,* and K. J. Friston*

*Wellcome Department of Cognitive Neurology, Institute of Neurology, and †Institute of Cognitive Neuroscience and Department of Psychology, University College London, Queen Square, London WC1N 3BG, United Kingdom

We introduce a new method for detecting differences in the latency of blood oxygenation level-dependent (BOLD) responses to brief events within the context of the General Linear Model. Using a first-order Taylor approximation in terms of the temporal derivative of a canonical hemodynamic response function, statistical parametric maps of differential latencies were estimated via the ratio of derivative to canonical parameter estimates. This method was applied to two example datasets: comparison of words versus nonwords in a lexical decision task and initial versus repeated presentations of faces in a fame-judgment task. Tests across subjects revealed both magnitude and latency differences within several brain regions. This approach offers a computationally efficient means of detecting BOLD latency differences over the whole brain. Precise characterization of the hemodynamic latency and its interpretation in terms of underlying neural differences remain problematic, however. © 2002 Elsevier Science

Key Words: event-related; fMRI; word; nonword; faces; BOLD; impulse; latency; deactivations.

Several authors have argued that analysis of the latency (as well as the magnitude) of the blood oxygenation level-dependent (BOLD) impulse response may prove informative with regard to the neural/synaptic activity following brief stimulation (e.g., Menon *et al.,* 1998; Kruggel and von Cramon, 1999; Miezin *et al.,* 2000). The present work introduces a new whole-brain statistical technique for testing differences in the latency of the BOLD impulse response function within the context of the General Linear Model (GLM). We use data from a lexical decision task and a face fame-judgment task to demonstrate the ability of this technique to detect latency differences within brain regions between two classes of stimuli: words versus nonwords

and initial versus repeated presentations of famous faces.

We rarely know the precise shape of the BOLD impulse response for a given brain region, but we can make an informed guess in light of knowledge of the canonical hemodynamic response function (HRF) derived from previous studies. To allow for some deviations about this canonical form, Friston *et al.* (1998) added further response functions derived from a first-order multivariate Taylor expansion of the canonical HRF. These functions included the partial derivative with respect to time (temporal derivative) and partial derivative with respect to duration (dispersion derivative). The inclusion of this set of "basis" functions within the General Linear Model allows estimation of the contribution of each basis function (its parameter estimate), linear combination of which allows calculation of the mean and standard error of the best-fitting event-related response. Friston *et al.* (1998) also proposed that tests of differences in the latency of responses can be derived from knowledge of the standard error of the fitted response. The present work takes this proposal further by explicitly estimating response latency (relative to the canonical HRF) via the ratio of temporal derivative to canonical parameter estimates. A preliminary application of this proposal was reported by Henson *et al.* (1999a), and a related, more general proposal has been made recently by Liao *et al.* (2001).

The canonical response function and its temporal derivative are shown in Fig. 1a. Assume that the real event-related BOLD response, $F(t)$, as a function of poststimulus time, t, is delayed relative to the canonical response, $f(t)$, by a small amount dt, such that

$$F(t) = \alpha f(t + dt),$$

where α is a scaling factor. With a first-order Taylor expansion of the canonical response,

$$f(t + dt) \sim f(t) + f'(t)dt,$$

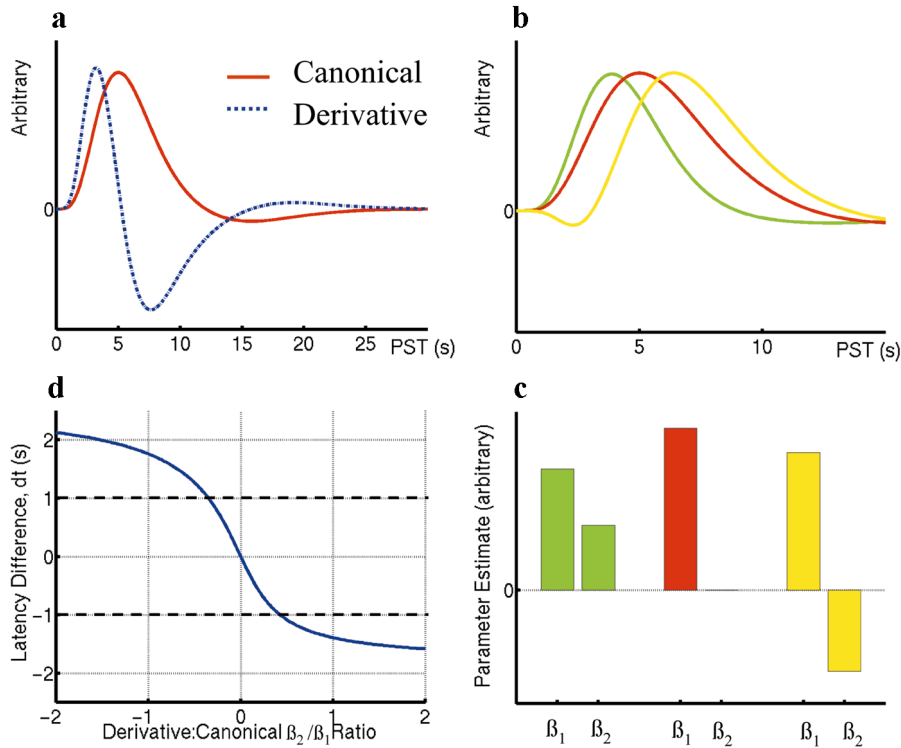¹ To whom correspondence should be addressed. E-mail: r.henson@ucl.ac.uk.

FIG. 1. (a) The canonical hemodynamic response function (HRF) and its temporal derivative plotted against poststimulus time (PST). The temporal derivative used in the present study is the finite difference between the canonical HRF and the same HRF shifted 1 s later in time. This difference function is then orthogonalized with respect to the canonical HRF to remove the small correlations introduced by the finite approximation to the continuous derivative. (b) The canonical HRF (red) together with HRFs shifted 1 s earlier (green) or later (yellow) in time. (c) Parameter estimates for canonical (β_1) and derivative (β_2) associated with HRFs in b. (d) The relationship between the latency difference relative to the canonical HRF (dt) and the ratio of derivative:canonical parameter estimates (β_2/β_1).

where $f'(t)$ is the temporal derivative, then

$$F(t) \sim \alpha f(t) + \alpha \, dt f'(t).$$

If the canonical response and its temporal derivative are used as two basis functions in a GLM, the parameter estimates, β_1 and β_2, derived from the least-means-squares solution of

$$F(t) = \beta_1 f(t) + \beta_2 f'(t)$$

are then such that

$$\alpha \sim \beta_1, \quad dt \sim \beta_2/\beta_1.$$

In other words, the magnitude of the BOLD response, α, is estimated by the canonical parameter estimate, and the latency of the BOLD response, dt (relative to the canonical response) is estimated by the ratio of derivative to canonical parameter estimates (provided dt is small relative to the time constants of the canonical HRF; see below).

Response functions shifted ± 1 s relative to the canonical are shown in Fig. 1b. The corresponding parameters for the canonical function (β_1) and its tempo-

FIG. 2. Regions showing magnitude effects for (i) main effect of word/nonword presentation versus baseline (top row, increases in red, decreases in blue) and (ii) differential responses to words and nonwords (middle row, greater responses to words in red, greater responses to nonwords in blue). Effects are rendered on a canonical normalized brain (with cerebellum artificially removed), in which color density decreases with distance from cortical surface. Below are plots for region (a) in left temporoparietal cortex ($x = -51$; $y = -57$; $z = +24$) and region (b) in left temporo-occipital cortex ($x = -48$; $y = -69$; $z = -6$). The left plot for each region shows the parameter estimates for the canonical (c) and derivative (d) for words (W) and nonwords (N). The right plot for each region shows the best fitting event-related response (solid lines, the sum of the two basis functions, weighted by their parameter estimates) and adjusted data binned every 2 s (dotted lines) as a function of poststimulus time (PST). Error bars refer to the standard error of parameter estimates/adjusted data across subjects (the standard error of the difference in word–nonword parameter estimates is inset).

FIG. 3. Regions showing latency effects (i) relative to canonical response (top row, delayed relative to canonical in blue) and (ii) differing for words and nonwords (middle row, delayed for words in red, delayed for nonwords in blue). Plots are for region (a) in right intraparietal sulcus ($x = +33$; $y = -48$; $z = +39$) and region (b) in left motor cortex ($x = -39$; $y = -12$; $z = +54$). See Fig. 2 legend for more details.
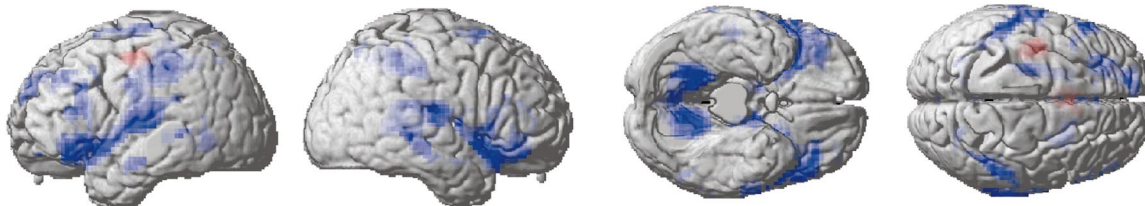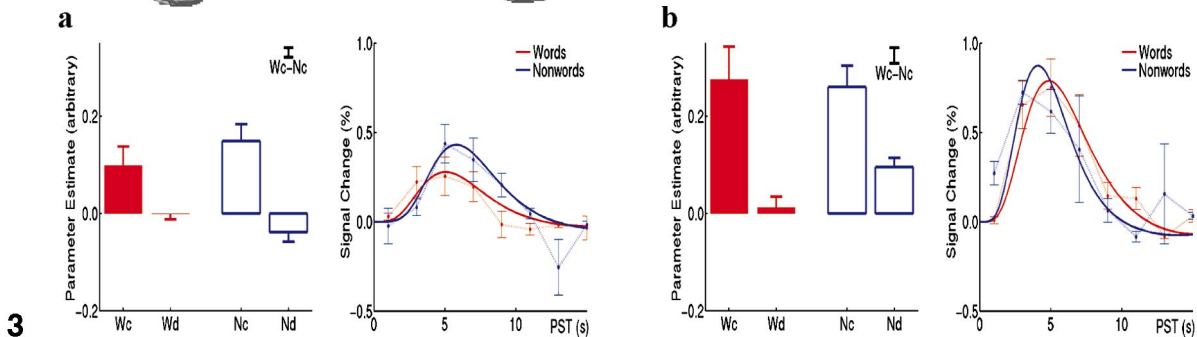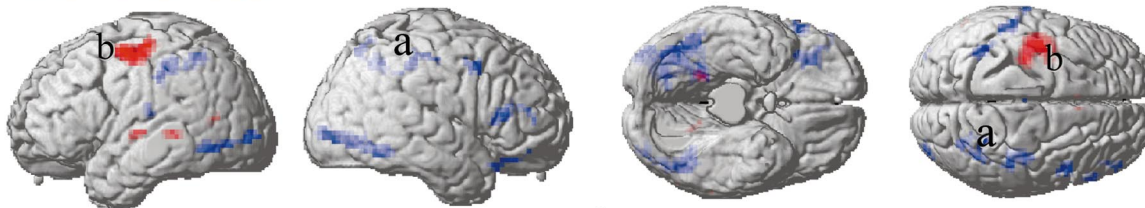
Words + Nonwords

Words − Nonwords

a

b

2

Words + Nonwords

Words − Nonwords

a

b

3
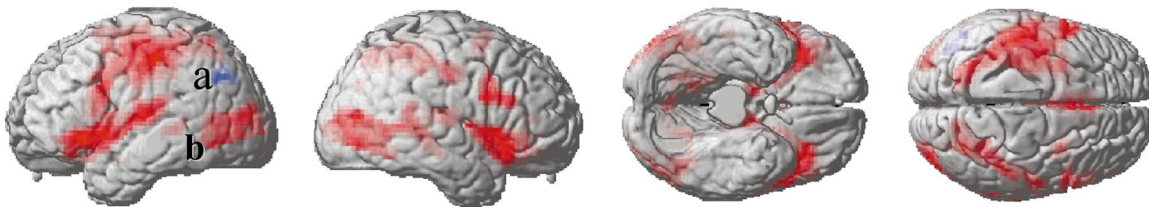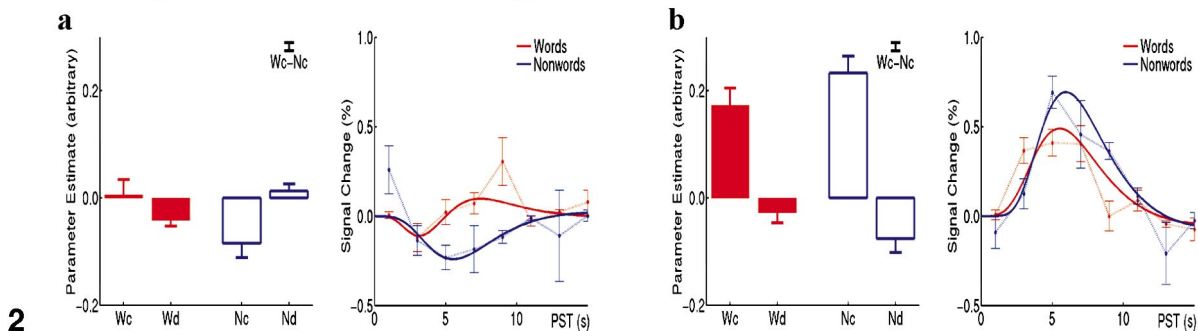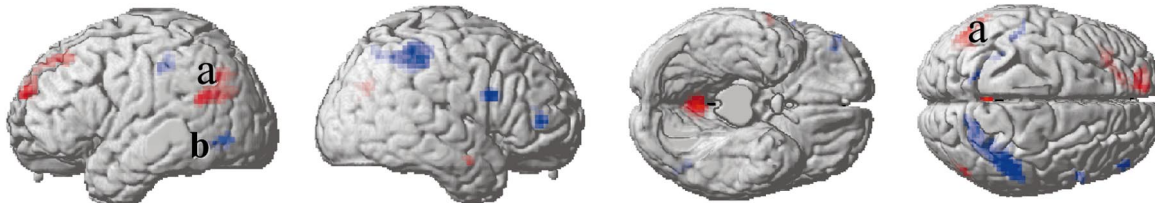
ral derivative (β_2) are shown in Fig. 1c. For responses earlier than the canonical, the derivative parameter is positive; for responses later than the canonical, the derivative parameter is negative.

The value of dt in seconds is plotted against the ratio of derivative:canonical parameter estimates in Fig. 1d. Positive ratios (negative values of dt) correspond to responses occurring earlier than the canonical HRF. The range of latency differences is approximately ± 2 s (given that fitted responses cannot peak earlier than the peak of the temporal derivative). The relationship between dt and the derivative:canonical ratio is almost linear for latency differences between ± 1 s, with a gradient of approximately -2.58 s.

An image of response latency can be created by transforming the derivative:canonical ratio for each voxel with a function resembling that in Fig. 1d. This function can be approximated by a sigmoidal logistic function,

$$2C/(1 + \exp(D\beta_2/\beta_1)) - C,$$

where $C = 1.78$ and $D = 3.10$ (root mean square error of nonlinear fit $= 0.18$). Statistical tests of latency differences can then be performed by comparing one such latency image for each event type across subjects, allowing construction of a statistical parameteric map (SPM) of the significance of latency differences across all voxels.

In summary, for small latency shifts, a first-order approximation is sufficient for estimating a linearly parameterized response. The ratio of the parameters for the HRF and its temporal derivative represent an estimate of that latency (and the same logic can be applied to other partial derivatives of a response function, such as the dispersion derivative (Friston *et al.,* 1998), which might be used to estimate the relative duration of BOLD impulse responses). However, this ratio is a nonlinear function of the actual latency by virtue of the fact that the parameterization ignores high-order terms. The above procedure can thus be seen as linearizing a nonlinear parameter estimation problem, solving using ordinary least squares, and applying a nonlinearity post hoc to account for higher order terms discounted during the estimation. The nonlinear transformation also renders the distribution of estimated latencies more Gaussian than the original ratio estimator (and we can appeal to the robustness of parameteric statistics to mild deviations from normality). The transformation is quick to perform and can be inserted after (subject-specific) parameter estimation, but before (multisubject) analyses over those parameters, maintaining a formal link with conventional two-stage, mixed-effects analyses used in fMRI (see Methods).

METHODS

Lexical Decision Dataset

This dataset comes from two sessions of a lexical decision task (which comprised the encoding phase of a memory experiment reported elsewhere; Henson *et al.,* 1999b). Previous blocked designs have revealed activations for the lexical decision task relative to a baseline task (e.g., Price *et al.,* 1994; Rumsey *et al.,* 1997), but could not distinguish differences in activation for the words and nonwords themselves.

The stimuli comprised two sets of 90 5-letter nouns with a Kucera–Francis written frequency of 10–100, selected from the MRC Psycholinguistics Database (http://www.psy.uwa.edu.au/uwa_mrc.htm). One set was used in each of the two 12-min sessions. For a random 30 words of each set, the letters were randomly permuted in order to create 5-letter nonwords. These nonwords were not therefore restricted to pseudowords (though some were pronounceable by virtue of the vowels). The 2:1 ratio of words to nonwords was selected on the basis of the subsequent memory experiment (Henson *et al.,* 1999b). Words and nonwords were displayed for 1 s, interspersed with 7 s of fixation (SOA of 8 s). The words were presented in a pseudorandom order, with the constraint that a nonword occurred every third stimulus on average, in order to improve sensitivity to the difference between words and nonwords (Josephs and Henson, 1999). Subjects indicated their lexical decision with one of two fingers of their right hand, the order of finger assignment being counterbalanced across subjects. The capitalized words were presented in 24-point Helvetica font on a Macintosh computer, projected onto a mirror approximately 300 mm above the subject in the MRI scanner. The horizontal visual angle subtended by the stimuli was approximately 4°.

Face Repetition Dataset

This dataset is a subset of that acquired during two different memory tasks performed on repeated presentations of famous and nonfamous faces (Henson *et al.,* 2001b). We restrict analysis here to correct fame decisions for initial versus repeated presentations of the famous faces during the fame-judgment task. Stimulus repetition tends to decrease decision times in such "implicit" memory tasks, which makes these data interesting in the context of repetition priming (Henson *et al.,* 2000).

Famous and nonfamous (unfamiliar) gray-scale faces were taken from the set created by Gorno-Tempini *et al.* (1998). A random sequence of two presentations of each face was created for each subject. The faces were presented for 500 ms against a baseline of an oval checkerboard present throughout the interstimulus interval, with a stochastic distribution of SOAs deter-

mined by a minimal SOA of 4.5 s and randomly inter-mixed null events with probability 1/3 (Josephs and Henson, 1999). Each stimulus was presented on a mirror above the participant, subtending a visual angle of approximately 10°. Each subject was scanned during two counterbalanced 20-min sessions, though only the session involving the fame-judgment task is considered here. Subjects were instructed to press one of two keys with either the index or the middle finger of their right hand to indicate whether or not each face belonged to a famous person, regardless of whether they had seen it before the experiment. The assignment of keys was counterbalanced across subjects.

Subjects

Different groups of 12 right-handed volunteers gave informed consent to participate in each experiment (ages 22–34 in the lexical experiment; ages 22–42 in the face experiment; 6 males in each group). All volunteers reported themselves to be in good health with no history of neurological illness.

Scanning Parameters

A 2-T Vision system (Siemens, Erlangen, Germany) was used to acquire T2*-weighted transverse echoplanar images (EPI) (64×64 pixels, TE = 40 ms) with BOLD contrast in a descending direction, positioned to cover the whole brain except cerebellum. In the lexical experiment, a total of 245 volumes of 34 2.5-mm-thick axial slices of 5×5-mm^2 pixels positioned every 3 mm were collected per session with an effective repetition time (TR) of 3 s/volume. In the face experiment, a total of 356 volumes of 24 3-mm-thick axial slices of 3×3-mm^2 pixels positioned every 4.5 mm were collected with an effective TR of 2 s/volume. The first 5 volumes in each session were discarded to allow for T1 equilibration effects. For the lexical experiment, the 8:3 ratio of SOA to TR ensured an effective sampling rate of the impulse response over trials of 1 Hz; for the face experiment, the 4.5:2 ratio ensured an effective sampling rate of 2 Hz. Because of their uneven proportions and pseudorandom ordering, the poststimulus sampling was not quite uniform for words and nonwords in the lexical experiment; the sampling for initial and repeated presentations of faces in the face experiment was uniform.

Data Analysis

Data were analyzed using SPM99 (Wellcome Department of Cognitive Neurology, London, UK; Friston *et al.*, 1995). All volumes were realigned spatially to the first volume, and the signal in each slice was then realigned temporally to that measured in the middle slice using a sinc interpolation. Note that one could also use the temporal derivative of a hemodynamic response function to allow for different slice timings (Henson *et al.*, 1999c): The temporal derivative of the canonical HRF used here, for example, could capture slice timing differences of approximately ± 1 s, i.e., TRs less than 2 s. This would obviate the need to realign the data temporally, which may alias frequencies above the sampling limit. Little power was expected above the sampling limits in the present experiments, however (with their combination of short TRs and longer relative SOAs), and temporal realignment was desirable to increase the likelihood that any latency differences between event types were close to the linear range in Fig. 1D. (Temporal realignment would also be particularly important if one wanted to make inferences regarding latency across slices.)

The resliced volumes were normalized to a standard EPI template based on the MNI reference brain (Ashburner and Friston, 1999) in Talairach space (Talairach and Tournoux, 1988). The normalized images of $3 \times 3 \times 3$-mm^3 voxels were smoothed with an 8-mm FWHM isotropic Gaussian kernel. Treating the volumes as a time series, the data were high-pass filtered to 1/60 Hz in the lexical experiment and 1/120 Hz for the face experiment (owing to the larger range of SOAs in the latter). The smoothed images were scaled to a grand mean of 100 over all voxels and scans within a session (note that this session-wide scaling does not have the caveats associated with scan-specific global scaling; Aguirre *et al.*, 1998).

For the lexical experiment, three event-types were defined: (1) correct decisions to words, (2) correct decisions to nonwords, and (3) incorrect decisions to words or nonwords (the latter, though rare, were modeled separately to capture additional error-related variability). For the face experiment, there were eight basic event types, derived from correct and incorrect responses to initial and repeated presentations of famous and nonfamous faces. Correct responses were further divided into those matched and unmatched for time within the session. This matching was performed to rule out the potential time confound resulting from the fact that repeated presentations occurred later, on average, than initial presentations (see Henson *et al.*, 2001b, for further details). The two event types of interest—initial versus repeated presentations of famous faces—were therefore controlled for nonspecific time effects.

Each analysis was performed in a two-stage, mixed-effects procedure (Holmes and Friston, 1998). In the first stage, the BOLD response for each event type was modeled with the canonical HRF and its temporal derivative (see the introduction). These functions were convolved with an event train of delta functions at each stimulus onset and downsampled every TR (at the midscan time point) to create covariates in a GLM, together with a constant term for each session. Parameter estimates for each covariate were calculated from
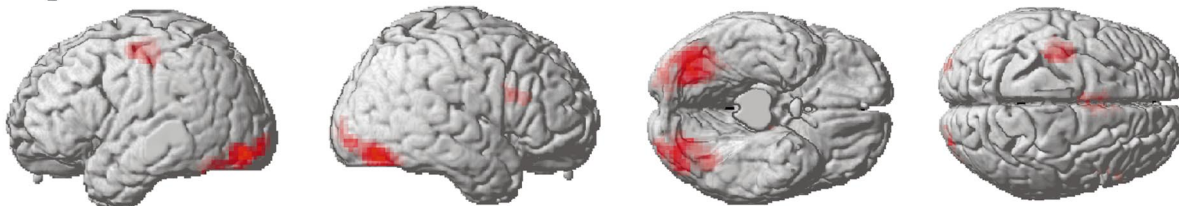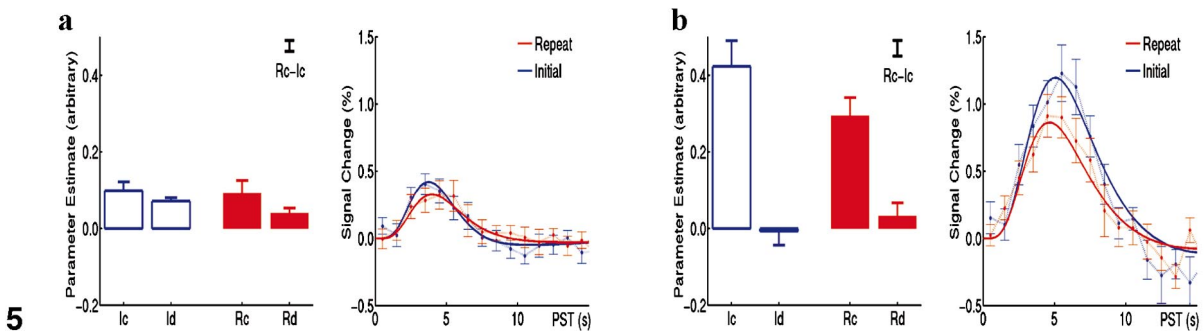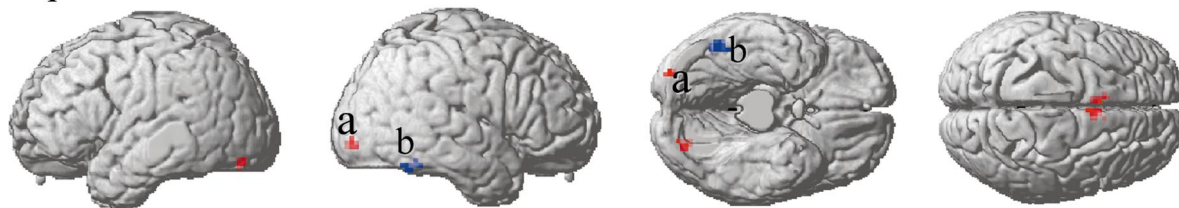
Repeat + Initial

Repeat - Initial

4

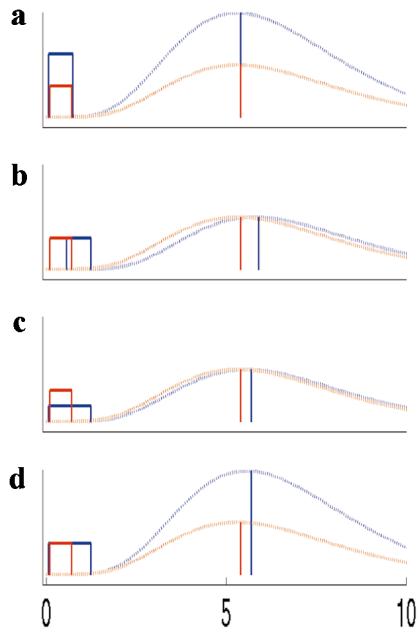Repeat + Initial

Repeat - Initial

5
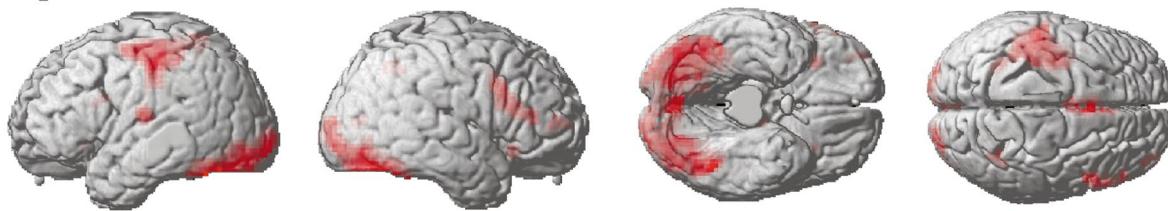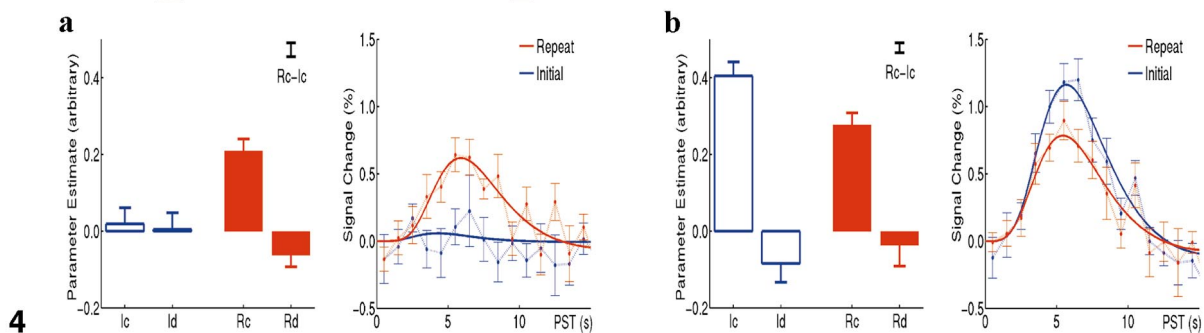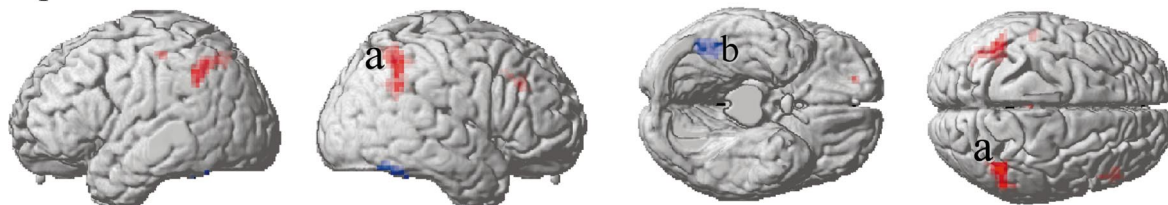
FIG. 6. BOLD responses (dotted lines) predicted from linear convolution of short bursts of neural/synaptic activity (rectangular functions) with the canonical HRF as a function of poststimulus time (in seconds) for two event types to illustrate effects of differences in (a) magnitude, (b) onset, and (c/d) duration of neural activity. The integrated neural activity is equated in c, whereas the peak neural activity is equated in d. Solid lines illustrate magnitude and latency of the peak BOLD response.

the least-mean-squares fit of the model to the time series. Images of the parameter estimates for the canonical and derivative covariates were created by subject-specific contrasts (collapsing across sessions within subjects). These "summary statistic" images comprised the data for a second stage of repeated-measures analyses (Frison and Pocock, 1992), treating subjects as a random variable. Pairwise one-tailed contrasts on the canonical parameter images allowed t tests on differences in the magnitude of event-related responses (t values were subsequently transformed to Z values). Two contrasts were tested: the main effect of the two event types (versus the interstimulus baseline) and the differential effect between the two types.

Pairwise tests on the differences in derivative parameter estimates cannot be directly interpreted as latency differences, because the effect of the temporal derivative depends on the canonical parameter esti-

mate (in both sign and magnitude). Larger responses (bigger canonical parameter estimates) require greater contributions of the temporal derivative to shift them forward or backward in time. Rather, derivative parameter estimates must be normalized by the canonical parameter estimates. "Latency maps" of this derivative:canonical ratio at each voxel, after a sigmoidal transformation (see the introduction), were smoothed by an 8-mm FWHM isotropic Gaussian kernel and entered into paired t tests to produce SPMs of latency differences.

To restrict analysis to voxels in which the canonical HRF provided a reasonable fit to the data, the second-stage SPMs were masked with voxels that survived $P < 0.05$ corrected in reduced F tests (separable across event types, pooled across subjects) from the first-stage SPMs. These tests identify voxels for which a significant proportion of the total variability in the evoked response to one or more event types was captured by the canonical HRF. Voxels not surviving these tests were therefore excluded from the analysis (since the present attribution of "magnitude" and "latency" is meaningful only in the context of impulse responses resembling the canonical form). Given that the primary concern of the present analyses was methodological rather than neuroscientific, more liberal uncorrected thresholds were used for the second-stage SPMs: The canonical SPMs were thresholded at $P < 0.005$ and the latency SPMs at $P < 0.025$ (corresponding to two-tailed P values of 0.01 or 0.05, respectively, since both directions were tested). Only regions comprising at least 10 contiguous voxels were reported. The maxima of such regions were localized as best as possible to the systems of Talairach and Tournoux (1988) and Brodmann (1909).

RESULTS

Lexical Decision Dataset

Behavioral Results

The means of the error rates and median correct reaction times across subjects were 0.03 and 931 ms, respectively, for words and 0.07 and 998 ms for non-words. The reaction times did not differ significantly, $F(1,11) = 2.21$, MSE $= 27,202$, $P > 0.16$.

FIG. 4. Regions showing magnitude effects for (i) main effect of face presentation versus baseline (top row, increases in red) and (ii) differential responses to initial and repeated face presentations (middle row, greater responses to repeated presentations in red, greater responses to initial presentations in blue). Below are plots for region (a) in right parietal cortex ($x = +45$; $y = -51$; $z = +54$) and region (b) in right lateral fusiform cortex ($x = +45$; $y = -48$; $z = -33$). See Fig. 2 legend for more details.

FIG. 5. Regions showing latency effects (i) relative to canonical response (top row, earlier than canonical in red, later than canonical in blue) and (ii) differing for initial and repeated face presentations (middle row, delayed for repeated presentations in red, delayed for initial presentations in blue). Plots are for region (a) in right inferior occipital cortex ($x = +27$; $y = -87$; $z = -6$) and region (b) in right lateral fusiform cortex ($x = +48$; $y = -54$; $z = -24$). See Fig. 2 legend for more details.

TABLE 1

Maxima within Regions Showing Word–Nonword Differences in Canonical Response Magnitude

Region of activation	Left/ right	Brodmann area	Volume (cm^3)	Talairach coordinates			Z value
				x	y	z	
(A) Words > nonwords							
Anterior medial frontal	L	9	3.97	−9	60	30	3.78
				−33	33	48	3.57
Anterior middle temporal	R	21	0.32	66	−3	−21	3.59
Cingulate	L	24	0.59	−12	−9	42	3.58
Posterior cingulate/precuneus	B	31/23	5.86	0	−57	18	3.86
Temporoparietal	**L**	**39**	**4.59**	**−51**	**−57**	**24**	**3.87**
				−42	−72	42	3.21
	R	39	0.54	54	−69	30	3.46
(B) Words < nonwords							
Inferior frontal	R	45/46	1.08	45	45	6	4.82
Inferior frontal	R	6/44	1.67	54	12	24	3.80
Superior/inferior parietal	L	7/40	2.59	−51	−27	42	3.51
				−27	−48	48	3.24
	R	7/40	11.53	45	−36	45	4.30
				24	−60	51	4.25
Temporo-occipital	**L**	**19/37**	**0.32**	**−48**	**−69**	**−6**	**2.94**

Note. L, left; R, right; B, bilateral; responses of regions in bold are plotted in Fig. 2.

Magnitude of Canonical Response

Words and nonwords relative to baseline. Extensive activations were seen for the main effect of word/nonword presentation versus baseline (top row, Fig. 2). These included regions in lateral temporo-occipital cortex, lingual gyri, intraparietal sulci, lateral fissures and insulae, pre- and postcentral (motor) cortex on the left, precentral sulci, posterior inferior prefrontal cortex, and bilateral medial frontal/anterior cingulate cortex. A deactivation was seen in a left temporoparietal region (see below).

Words greater than nonwords. Several regions showed differential responses to words and nonwords (middle row, Fig. 2). Greater responses to words were seen in bilateral temporoparietal cortex, posterior cingulate/precuneus, and dorsal cingulate, as well as left anterior medial frontal and right anterior middle temporal cortex (Table 1A). The left temporoparietal region, which included the ascending ramus of the superior temporal sulcus and aspects of the angular gyrus (BA 39), has previously been associated with semantic processing (see, for example, Price, 2000). This region showed a deactivation relative to baseline for nonwords (region "a" in Fig. 2), post hoc $t(12) = 3.15$, $P < 0.005$ (but not for words, $t(12) = 0.24$). This BOLD signal decrease may reflect greater neural activity during the baseline period (intertrial fixation) than immediately following nonword presentation. Such transient decreases have been shown to produce negative BOLD responses similar to inverted normal impulse responses (though differing in the postpeak undershoot; Fransson *et al.,* 1999).

Nonwords greater than words. Regions showing greater responses to nonwords than to words included extensive parietal regions (mainly within the intraparietal sulcus, particularly on the right, but extending into both supramarginal gyri), two regions in right inferior frontal cortex, and one in left temporo-occipital cortex (Table 1B). The latter region (region "b" in Fig. 2), most probably in anterior occipital sulcus (BA 19/37), has previously been associated with phonological retrieval (Price, 1998). This region showed activations for both words and nonwords relative to baseline (post hoc $t(12) > 5.4$, $P < 0.001$), but relatively greater activation for nonwords. This activation also peaked later for nonwords than for words (see also Fig. 3).

Latency of Canonical Response

Words and nonwords relative to canonical. The responses of many of the regions that showed a main effect of word/nonword presentation versus baseline were delayed relative to the canonical response (top row, Fig. 3), the main exceptions being the bilateral temporo-occipital, left motor, and bilateral medial frontal (SMA) regions. The estimated latency relative to the canonical (*dt*), averaged over the voxels with delayed responses, was 0.41 s (ranging between 0.13 and 0.78 s). Two regions, in left motor cortex and bilateral SMA, showed earlier responses than the canonical.

The finding that BOLD responses in frontotemporal

TABLE 2

Maxima within Regions Showing Word–Nonword Differences in Canonical Response Latency

Region of activation	Left/ right	Brodmann area	Volume (cm^3)	Talairach coordinates			Z value
				x	y	z	
(A) Words > nonwords							
Medial frontal	B	6/8	0.46	3	18	63	2.76
Precentral gyrus	**L**	**4/6**	**6.45**	**−39**	**−12**	**54**	**3.34**
				−30	−21	60	3.26
Superior temporal gyrus	L	21	0.70	−57	−12	0	2.90
Superior temporal sulcus	L	21	0.76	−60	−39	0	3.03
Lingual/parahippocampal	L	19/30	0.27	−15	−51	6	2.37
	R	19/30	0.49	21	−45	3	2.85
Posterior insula	L	—	1.24	−39	−24	12	2.71
Superior temporal sulcus	L	39	0.27	−54	−66	12	2.29
(B) Words < nonwords							
Inferior frontal	R	47/11	1.92	30	33	−18	3.57
Inferior frontal	R	45/46	0.97	57	27	12	3.32
Middle frontal	R	6/9	0.89	51	12	42	3.56
Middle frontal	R	46	0.78	39	30	15	2.58
	R	—	1.43	36	−6	−9	3.63
Cingulate	B	24	0.46	0	−24	45	2.48
Inferior parietal	L	40	0.32	−66	−21	18	2.80
Inferior parietal	L	40	1.51	−54	−33	48	2.69
Superior parietal	L	7	1.11	−33	−54	51	2.80
Inferior/superior parietal	**R**	**7/40**	**5.80**	**33**	**−48**	**39**	**3.85**
				33	−69	54	3.22
Temporo-occipital	L	19/37	3.00	−39	−54	−12	3.03
				−33	−87	−3	2.94
	R	19/37	8.61	36	−63	−6	4.28
				21	−72	−9	2.99

Note. L, left; R, right; B, bilateral; >, later than; <, earlier than; responses of regions in bold are plotted in Fig. 3.

regions were delayed relative to the canonical, whereas those in motor and SMA regions were earlier than the canonical, may appear counterintuitive, given that the motor and SMA responses presumably reflect relatively later components of responding (e.g., right finger depression). However, the possible differences in vasculature (i.e., precise mapping from neural to BOLD responses) render such across-region comparisons uninterpretable (see also Miezin *et al.,* 2000, who found that responses in motor cortex could peak earlier than those in visual cortex during a visual–motor response task).

Words earlier than nonwords. Several regions showed differential latencies to words and nonwords (middle row, Fig. 3). Regions showing shorter latencies to words included bilateral temporo-occipital and intraparietal cortex, bilateral insulae, and right inferior frontal cortex (Table 2B). The right parietal region (region "a" in Fig. 3), which has previously been activated for phonological relative to semantic judgments (Mummery *et al.,* 1998), showed both a greater and a delayed activation for nonwords than for words (see also Fig. 2). The relative delay for nonwords was seen

in 9 of the 12 subjects. The mean estimated latencies in this region (*dt*) were −0.42 and 0.65 s for words and nonwords, respectively, a difference of −1.07 s. This differs from the peak-to-peak difference of −0.75 s in the fitted responses of Fig. 3 because of the averaging of a nonlinear transformation (in the former case), rather than the averaging of a linear transformation (in the latter case).

Nonwords earlier than words. Regions showing earlier responses to nonwords included left motor cortex (precentral gyrus), medial frontal cortex, and several regions in the left superior temporal sulcus/gyrus (Table 2A). Responses in the left motor region (region "b" in Fig. 3), most probably reflecting the right finger presses, were of approximately equal magnitude for words and nonwords, simply delayed for words relative to nonwords. The relative delay for words was seen in 11 of the 12 subjects. The mean estimated latencies relative to the canonical response were −0.04 and −1.00 s for words and nonwords, respectively, a difference of 0.96 s (cf. the peak-to-peak difference of 0.76 s in Fig. 3). This longer relative latency for words in left motor cortex (which can be seen in the data too, Fig. 3)

TABLE 3

Maxima within Regions Showing Differences in Canonical Response Magnitude to Initial
and Repeated Face Presentations

Region of activation	Left/right	Brodmann area	Volume (cm^3)	Talairach coordinates			Z value
				x	y	z	
(A) Repeated > initial							
Anterior frontal	R	10	0.32	21	51	0	2.92
Inferior frontal	R	9/46	0.68	54	30	30	3.38
Posterior cingulate	B	7/31	0.29	0	−27	48	3.36
Postcentral	L	1/2/3	0.35	−51	−27	54	3.06
Superior/inferior parietal	L	7/40	3.92	−45	−51	36	3.87
				−30	−72	54	3.06
	R	**7/40**	**5.83**	**45**	**−51**	**54**	**3.67**
				45	−48	30	3.98
(B) Repeated < initial							
Lateral fusiform	**R**	**37**	**2.51**	**45**	**−48**	**−33**	**3.54**

Note. L, left; R, right; B, bilateral; responses of regions in bold are plotted in Fig. 4.

would not be expected from the reaction times, which did not differ significantly (and if anything, were slightly longer for nonwords). Resolving this apparent paradox is not possible without more direct measures of motor cortex activity (e.g., lateralized readiness potentials, which might diverge earlier for nonwords than for words, even though the motor execution latency might not differ).

Face Repetition Dataset

Behavioral Results

The means of the error rates and median correct reaction times were 0.18 and 945 ms, respectively, for initial presentations of famous faces and 0.14 and 773 ms for repeated presentations. Reaction times to repeated presentations were significantly faster than to initial presentations, $F(1,11) = 20.35$, MSE = 8732, $P < 0.001$.

Magnitude of Canonical Response

Face presentations relative to baseline. Activations for the main effect of face presentation versus baseline (top row, Fig. 4) included an extensive region covering bilateral occipitotemporal and fusiform cortex, a region including left pre- and postcentral gyri (including motor cortex), and regions in left inferior central sulcus, medial frontal cortex in cingulate sulcus (and extending into SMA), left thalamus, bilateral anterior insulae, bilateral precentral sulci, right intraparietal sulcus, right inferior frontal gyrus, and right anterior frontal cortex. No region showed a deactivation relative to baseline.

Repeat greater than initial presentations. A num-

ber of regions showed differential responses to initial and repeated presentations of faces (middle row, Fig. 4). Most showed greater responses to repeated presentations (Table 3), including a right anterior frontal region, right inferior frontal sulcus, postcentral gyrus, posterior cingulate sulcus, and bilateral intraparietal regions. The parietal regions in particular have previously been associated with "old–new" effects in episodic recognition (see Rugg and Henson, 2001). Indeed, the right intraparietal region (region "a" in Fig. 4) appeared to respond only to repeated presentations, post hoc $t(12) = 6.89$, $P < 0.001$ for repeated, and $t(12) = 0.44$ for initial, presentations.

Initial greater than repeat presentations. Only one region showed a greater response to initial than to repeated presentations, in right lateral fusiform (BA 37), region "b" in Fig. 4. This region responded to both initial and repeated presentations (post hoc $t(12) > 9.06$, $P < 0.001$), but less so on repeated presentations. A repetition-related reduction in the response to familiar faces in this region, which is close to the fusiform face area defined by Kanwisher *et al.* (1997), has been shown previously (Henson *et al.,* 2000) and may reflect the behavioral phenomenon of repetition priming (e.g., Ellis *et al.,* 1990). This region also showed evidence of shorter latency for repeated than for initial presentations (see below).

Latency of Canonical Response

Face presentations relative to canonical. Several regions that showed a main effect of face presentation also showed a shorter latency than the canonical (top row, Fig. 5), namely bilateral inferior temporo-occipital and fusiform cortices, left motor cortex, and medial

TABLE 4

Maxima within Regions Showing Differences in Canonical Response Latency to Initial and Repeated Face Presentations

Region of activation	Left/ right	Brodmann area	Volume (cm³)	Talairach coordinates			Z value
				x	y	z	
(A) Repeated > initial							
Medial frontal	L	6/8	0.51	−6	12	54	2.74
	R	6/8	0.46	9	6	51	2.73
Inferior occipital	L	18	0.43	−24	−78	−18	2.42
	R	**18**	**0.30**	**27**	**−87**	**−6**	**2.42**
(B) Repeated < initial							
Lateral fusiform	**R**	**37**	**0.86**	**48**	**−54**	**−24**	**2.40**
				45	−45	−21	2.57

Note. L, left; R, right; B, bilateral; >, later than; <, earlier than; responses of regions in bold are plotted in Fig. 5.

frontal (SMA) regions. No region showed responses with a longer latency than the canonical.

Initial earlier than repeat presentations. A few regions showed differential latencies to initial and repeated face presentations (middle row, Fig. 5). Regions showing earlier responses to initial presentations occurred in bilateral posterior inferior occipital cortex and bilateral medial frontal cortices (Table 4A). The right inferior occipital region (region "a" in Fig. 5) showed a slightly smaller and later response to repetitions. The relative delay for repetitions was seen in 10 of the 12 subjects. The mean estimated latencies in this region (*dt*) were −1.44 and −0.36 s for initial and repeated faces, respectively, a mean difference of −1.08 s. This differed considerably from the value of only −0.25 s from the peak-to-peak difference in the fitted responses in Fig. 5. This discrepancy, owing to the nonlinear transform (see above), is likely to increase as the latency difference between the actual and the canonical response increases (e.g., beyond the near-linear region in Fig. 1D). In this case, the latencies relative to stimulus onset for the fitted responses to initial (3.88 s) and repeated (4.12 s) presentations were much earlier than for the canonical (5.01 s; see also top row, Fig. 5).

Repeat earlier than initial presentations. The only region showing an earlier response to repeats was the right lateral fusiform region, which also showed a smaller response magnitude to repeats (region "b" in Fig. 5). The mean estimated latencies were 0.09 and −0.58 s for initial and repeated presentations, respectively, a mean difference of 0.68 s (cf. a value of 0.37 s from the peak-to-peak difference in the fitted responses in Fig. 5). The shorter latency for repetitions was seen in 10 of the 12 subjects. This ability to detect a shorter latency to repeated face presentations in this region is consistent with nonlinear fitting of a response function

parametrized by onset latency, peak latency, and peak magnitude, as reported by Henson and Rugg (2001). These authors found a significant difference in median peak latency, but no significant difference in onset latency. This combination of a shorter peak latency and reduced peak magnitude is consistent with a shorter duration of underlying neural/synaptic activity following repetition (see Discussion).

DISCUSSION

The present study has demonstrated the feasibility of using the temporal derivative of a canonical hemodynamic response function to infer differences in BOLD impulse response latency. We were able to identify a number of regions that showed differences in both the magnitude and the latency of the event-related response to words and nonwords in a lexical decision task and to initial and repeated presentations of famous faces in a fame-judgment task. Such latency data may comprise a useful additional source of information with which to generate or contrast theories. We briefly consider these findings, before discussing the method in more detail.

Present Findings: Words versus Nonwords

Psychological interpretations of our word–nonword differences are limited because the stimuli were not well controlled (e.g., for word regularity, nonword pronunciability, syllabification, or word/nonword ratio, given that the primary interest was in subsequent memory for the words; Henson *et al.,* 1999b) and because any word–nonword differences in a lexical decision task could reflect the different decisions (acceptance versus rejection) rather than differences in the stimuli per se. Caution should also be exercised given

the liberal statistical thresholds used for illustrative purposes. Nonetheless, the general consistency between the present results and those of previous studies is reassurring. Regions showing greater responses to words were more extensive on the left and included posterior temporoparietal cortex, precuneus/posterior cingulate, and anterior medial prefrontal cortex. All of these regions have previously been associated with lexicosemantic processing: left temporoparietal region with semantic processing by both neuroimaging and neuropsychological studies (see Price, 2000, for review), precuneus/posterior cingulate in silent viewing of words versus false fonts (Price *et al.,* 1994), and left anterior medial prefrontal cortex for words versus nonwords in a feature detection task (Price *et al.,* 1996) and in a reading task (Rumsey *et al.,* 1997).

In the case of the left temporoparietal region, the word–nonword difference was in the context of a deactivation to nonwords (but not words) versus baseline. One possibility is that this deactivation reflects a transient cessation in neural activity that was occurring during the interstimulus interval. This interstimulus activity might reflect the demands of maintaining eye fixation between stimuli, for example (in which case, the interruption of fixation would need to be greater for nonwords than for words). Another possibility is that, when subjects are not engaged in an active task, they engage in free-flowing thought and this involves activity in semantic/conceptual processing areas (Binder *et al.,* 1999). According to the latter account, a stimulus-locked reduction in conceptual processing might be expected when the stimuli are nonwords, which have no semantic associations, but not words, which do.

Apart from a left temporo-occipital region, the regions that showed greater responses to nonwords than to words tended to be more extensive on the right. These included bilateral intraparietal sulci, extending to both inferior (supramarginal) and superior parietal gyri, and right inferior prefrontal cortex. These regions have previously been associated with phonological retrieval: the left temporo-occipital region is consistently activated during word retrieval tasks (see Price, 1998), and the parietal regions are more active when phonological judgments are contrasted with semantic judgments (Price *et al.,* 1997; Mummery *et al.,* 1998; see also Rumsey *et al.,* 1997).

While we did not have prior predictions for latency differences between words and nonwords, we illustrate one possible account of the pattern found in the left temporo-occipital region, which showed both a greater and a longer latency activation for nonwords than for words. Greater activation of this region for nonwords or pseudowords than for words has been reported previously in various tasks (Hagoort *et al.,* 1999; Rumsey *et al.,* 1997; Price *et al.,* 1996). It has been suggested that this increase in mean activation reflects the search for a lexical representation that has not been established by prior experience (Price *et al.,* 1996). In the context of competitive network models of visual word recognition (e.g., Seidenberg and McClelland, 1989), in which lexical units compete with one another, the delayed response to nonwords compared to words could reflect earlier resolution of this competition in the case of words compared to nonwords. Relatively prolonged neural activity for nonwords relative to words in the left temporo-occipital region would produce both the larger and the relatively delayed peaking of the BOLD response (see Fig. 6D and below).

Present Findings: Initial versus Repeated Faces

The differences between initial and repeated presentations of famous faces in the fame-judgment task are also consistent with previous findings. The enhanced response associated with repetition in regions such as right prefrontal, posterior cingulate, and bilateral parietal cortices has been associated with episodic retrieval (Rugg and Henson, 2001). The response of the right parietal region, which occurred only for repeated presentations, may reflect incidental recollection of the previous presentation of a face, for example.

Recent neuroimaging findings have also suggested that when a process is repeated (i.e., occurring on both first and second presentations), decreased responses are seen in brain regions associated with that processing (Schacter and Buckner, 1998). This is consistent with the right lateral fusiform region identified in the present experiment, which has previously been associated with face processing (Kanwisher *et al.,* 1997). In particular, if the process subserved by this region were recognition of individual faces, decreases would be expected for familiar (e.g., famous) faces, which would be recognized on both initial and repeated presentations, but not for unfamiliar faces, which might be recognized only after repeated presentations (Henson *et al.,* 2000).

Decreased hemodynamic responses following repetition may be a consequence of decreases in the mean firing rate of neuronal populations, analogous to the phenomenon of "response suppression" recorded from single cells in monkey inferior temporal cortex (Desimone, 1996). It has been suggested that these decreases also correlate with behavioral priming effects (Wiggs and Martin, 1998), which is consistent with the reaction time improvement associated with famous face repetition in the present fame-judgment task. However, the finding that both the magnitude and the latency of the fusiform response were reduced following repetition is also consistent with a shortened duration of neural/synaptic activity (see below and Henson and Rugg, 2001). This decreased duration could in fact occur without any change in the instantaneous level of neural activity (cf. Figs. 6a and 6d).

Advantages and Limitations of Present Method

There are several caveats associated with the present method. Foremost, care must be taken in extrapolating differences in BOLD latency to differences in the latency of underlying neural/synaptic activity. Assuming a linear convolution model (Friston *et al.,* 1998), a difference in BOLD latency can derive from a difference in the onset of neural activity (Fig. 6b) or a difference in the duration of simultaneously onsetting neural activity (Fig. 6c). These possibilities can, in principle, be distinguished by testing whether the difference arises in onset latency or peak latency: In Fig. 6B, the BOLD responses will differ equally in their onset and peak latency, whereas in Fig. 6c, the responses will differ in peak latency but not onset latency. The temporal derivative shifts the canonical response only earlier or later in time, however, with minimal change in its shape, and thus cannot distinguish these two types of latency shift. In the case of an equivalent level of neural activity that is simply sustained longer for one event type than for another, the BOLD response can differ in both latency and magnitude (Fig. 6d). The left temporo-occipital region, showing differences between words and nonwords, and the right fusiform region, showing differences between initial and repeated face presentations, both showed this pattern of combined magnitude and latency differences. Thus we cannot determine whether these BOLD differences reflect neural activity that is both stronger and delayed in one case than in the other or is simply more prolonged.

Several researchers have used models of the BOLD response that are explicitly parameterized by, for example, magnitude, latency, and duration (Kruggel and von Cramon, 1999; Miezin *et al.,* 2000; Henson and Rugg, 2001). Unlike the approach adopted here, these approaches are able, in principle, to distinguish BOLD onset latency from BOLD peak latency (potentially distinguishing cases b–d in Fig. 6) and hence allow more constrained inferences about magnitude, duration, and latency of underlying neural activity. However, these approaches require iterative search methods in order to fit the model to the data, which can be computationally expensive when performed for every voxel. Though our approach to detecting latency differences is less precise, it has the advantage of operating within the general linear model and therefore requiring considerably less computation in order to construct parametric maps of latency differences over the whole brain.

The model in Fig. 6 assumes a linear relationship between neural activity and the shape of the ensuing BOLD response. This is likely to be an oversimplification, given that significant nonlinearities, such as that between stimulus duration and BOLD magnitude, have been demonstrated (Boynton *et al.,* 1996; Glover,

1999; Vazquez and Noll, 1998). Such nonlinearities further complicate the inferences that can be made from BOLD response characteristics to underlying neural activity. For example, the model in Fig. 6 suggests that the magnitude and latency of the BOLD response are independent. It is possible that the coupling between neural activity and BOLD signal is such that stronger neural activity produces both a larger and a more delayed BOLD response (though Miezin *et al.,* 2000, failed to find a reliable correlation between magnitude and latency of BOLD responses across individuals, and several apparent dissociations between magnitude and latency were observed in various regions of the present study, e.g., Figs. 3 and 5).

In principle, the present approach could be used to make inferences about differential BOLD latencies across different brain regions (Thierry *et al.,* 1999), simply via maps of the mean latency, relative to the canonical response, for all voxels. While this approach may be useful for investigating different hemodynamic latencies across different brain regions, interpretation of such differences in terms of underlying neural/synaptic activity presents even greater difficulties than those already discussed, because they could reflect differences in the vascular dynamics of different brain regions (Lee *et al.,* 1995), rather than differences in the dynamics of cognitive processes.

The present approach assumes that the parameter estimates for the HRF and its temporal derivative can be estimated efficiently for each event type. Though the temporal derivative used here is orthogonal to the canonical HRF, meaning that the corresponding regressors for a given event type are minimally correlated (assuming minimal undersampling of the functions), correlations can arise between the regressors of different event types, which will reduce the estimation efficiency. If the events of one type consistently follow those of another by a small delay, for example, the correlation between their regressors will be high (e.g., the difference between the canonical regressors will be highly correlated with the regressors for their temporal derivatives), preventing accurate estimation of any latency differences. This reflects a general problem in estimating the shape of the BOLD impulse response, which, for short SOAs, benefits from random ordering of event types and a stochastic distribution of SOAs (Friston *et al.,* 1999; Josephs and Henson, 1999).

Another caveat is that, while the present approach may be suitable for detecting latency differences, it is not a precise method for quantifying those latencies. Because latency is estimated by the ratio of two parameters, which are themselves estimated, the estimation error in the ratio is high. This produces ratios that can vary considerably across subjects. The nonlinear sigmoidal transform renders these ratios more normal across subjects, but it also weakens the correspondence between transformed latency estimates and the laten-

cies in the mean fitted responses (which are linear functions of the canonical and derivative parameter estimates), as exemplified under Results. The present approach may thus be useful in conjunction with other methods, in which it functions as an exploratory, whole-brain analysis to identify regions of interest, which may then benefit from more precise but computationally expensive nonlinear methods (e.g., Kruggel and von Cramon, 1999; Miezin *et al.,* 2000; Henson and Rugg, 2001).

A final caveat of the present approach is that the derivative:canonical ratio has little meaning when the canonical response function does not approximate the real BOLD impulse response (which is why the present analyses were restricted to voxels in which the canonical response function accounted for significant variability; see Methods). Indeed, regions with impulse responses that are sufficiently noncanonical will not be identified at all. In the present study, the close match between the data and the fitted responses (in Figs. 2–5) confirms that the model fit is good (and in the case of the face fame-judgment data, we have found little evidence for other brain regions that responded to faces with a noncanonical BOLD response; Henson *et al.,* 2001a). If the data differed markedly from the fitted response, however, the present estimation of latency becomes ill defined. For example, a systematic deviation in the initial or subsequent undershoot of the BOLD impulse response, relative to the canonical HRF, might be partially captured by a nonzero temporal derivative parameter, despite an absence of any difference in peak/onset latency. Nonetheless, the logic of including the temporal derivative of an alternative hemodynamic response function in a first order Taylor approximation still holds (e.g., using a response function that includes greater initial or subsequent undershoots).

A similar caveat applies to differences in BOLD magnitude inferred from tests on the canonical parameter estimate alone: If one response is appreciably delayed with respect to the canonical, it will also have a smaller canonical parameter estimate. A trend to this effect can be seen in Fig. 1C, in which the canonical parameter estimate (β_1) is slightly reduced for responses that peak 1 s earlier or later than the canonical, even though those peaks are the same height (Fig. 1B). Though this reduction may be appreciable only for latency differences of more than 1 s, significant differences in the canonical parameter estimate can really be interpreted only as differences in response magnitude if there is no concomitant difference in response latency.

CONCLUSION

We have shown how the temporal derivative of a hemodynamic response function can allow an efficient means of creating whole-brain images of latency differences. The validity of this approach receives support from the statistical identification of a subset of brain regions for which the adjusted data showed latency differences between two event types. The interpretation of such latency differences must be qualified, however, by the difficulty in extrapolating from BOLD response characteristics to the characteristics of underlying neural activity.

REFERENCES

Aguirre, G. K., Zarahn, E., and D'Esposito, M. 1998. The inferential impact of global signal covariates in functional neuroimaging analyses. *NeuroImage* **8:** 302–306.

Ashburner, J., and Friston, K. 1999. Nonlinear spatial normalization using basis functions. *Hum. Brain Mapp.* **7:** 254–266.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Rao, S. M., and Cox, R. W. 1999. Conceptual processing demands during the conscious resting state: A functional MRI study. *J. Cognit. Neurosci.* **11:** 80–93.

Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. 1996. Linear systems analysis of functional magnetic resonance imaging data in human V1. *J. Neurosci.* **16:** 4207–4221.

Brodmann, K. 1909. *Vergleichende Lokalisationslehre der Grosshirnrinde in Ihren Prinzipien Dargelstellt auf Grund des Zellesbaues.* Leipzig: Barth.

Desimone, R. 1996. Neural mechanisms for visual memory and their role in attention. *Proc. Natl. Acad. Sci. USA* **93:** 13494–13499.

Ellis, A. W., Young, A. W., and Flude, B. M. 1990. Repetition priming and face processing: Priming occurs within the system that responds to the identity of a face. *Q. J. Exp. Psychol.* **42A:** 495–512.

Fransson, P., Kruger, G., Merboldt, K. D., and Frahm, J. 1999. MRI of functional deactivation: Temporal and spatial characteristics of oxygenation-sensitive responses in human visual cortex. *NeuroImage* **9:** 611–618.

Frison, L., and Pocock, S. J. 1992. Repeated measures in clinical trials: Analysis using mean summary statistics and its implications for design. *Stat. Med.* **11:** 1685–1704.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., and Turner, R. 1998. Event-related fMRI: Characterizing differential responses. *NeuroImage* **7:** 30–40.

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

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N. A., and Dale, A. 1999. Stochastic designs in event-related fMRI. *NeuroImage* **10:** 607–619.

Glover, G. H. 1999. Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage* **9:** 416–429.

Gorno-Tempini, M. L., Price, C. J., Josephs, O., Vandenberghe, R., Cappa, S. F., Kapur, N., Frackowiak, R. S., and Tempini, M. L. 1998. The neural systems sustaining face and proper-name processing. *Brain* **121:** 2103–2118.

Hagoort, P., Indefrey, P., Brown, C., Herzog, H., Steinmetz, H., and Seitz, R. J. 1999. The neural circuitry involved in the reading of

German words and pseudowords: A PET study. *J. Cognit. Neurosci.* **11:** 383–398.

Henson, R. N. A., Buechel, C., Josephs, O., and Friston, K. 1999c. The slice-timing problem in event-related fMRI. *NeuroImage* **9:** 125.

Henson, R. N. A., and Rugg, M. D. 2001. Effects of stimulus repetition on latency of the BOLD impulse response. *NeuroImage* **13:** 683.

Henson, R. N. A., Rugg, M. D., and Friston, K. 2001a. The choice of basis functions in event-related fMRI. *NeuroImage* **13:** 149.

Henson, R. N. A., Rugg, M. D., Shallice, T., Josephs, O., and Dolan, R. 1999b. Recollection and familiarity in recognition memory: An event-related fMRI study. *J. Neurosci.* **19:** 3962–3972.

Henson, R. N. A., Shallice, T., and Dolan, R. 2000. Neuroimaging evidence for dissociable forms of repetition priming. *Science* **287:** 1269–1272.

Henson, R. N. A., Shallice, T., Gorno-Tempini, M. L., and Dolan, R. 2001b. Face repetition effects in implicit and explicit memory tests as measured by fMRI. *Cerebral Cortex,* in press.

Henson, R. N. A., Shallice, T., Price, C., Dolan, R. J., Friston, K., and Turner, R. 1999a. Lexical decision: Differences in magnitude and onset as indexed by event-related fMRI. *NeuroImage* **9:** 1044.

Herbster, A. N., Mintun, M. A., Nebes, R. D., and Becker, J. T. 1997. Regional cerebral blood flow during word and nonword reading. *Hum. Brain Mapp.* **5:** 84–92.

Holmes, A. P., and Friston, K. J. 1998. Generalisability, random effects and population inference. *NeuroImage* **7:** 754.

Josephs, O., and Henson, R. N. A. 1999. Event-related fMRI: Modelling, inference and optimisation. *Philos. Transact. R. Soc. London* **354:** 1215–1228.

Kanwisher, N., McDermott, J., and Chun, M. M. 1997. The fusiform face area: A module in human extrastriate cortex specialised for face perception. *J. Neurosci.* **17:** 4302–4311.

Kruggel, F., and von Cramon, D. Y. 1999. Modeling the hemodynamic response in single-trial functional MRI experiments. *Magn. Reson. Med.* **42:** 787–797.

Lee, A. T., Glover, G. H., and Meyer, C. H. 1995. Discrimination of large venous vessels in time-course spiral blood-oxygenation-level-dependent magnetic-resonance functional imaging. *Magn. Reson. Med.* **33:** 745–754.

Liao, C. H., Worsley, K. J., Poline, J.-B., Duncan, G. H., and Evans, A. C. 2001. Estimating the delay of the hemodynamic response in fMRI data. Submitted for publication.

Menon, R. S., Luknowsky, D. C., and Gati, J. S. 1998. Mental chronometry using latency-resolved functional MRI. *Proc. Natl. Acad. Sci. USA* **95:** 10902–10907.

Miezin, F. M., Maccotta, L., Ollinger, J. M., Peterson, S. E., and Buckner, R. L. 2000. Characterizing the hemodynamic response: Effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *NeuroImage* **11:** 735–759.

Mummery, C. J., Patterson, K., Hodges, D., and Price, C. J. 1998. Organisation of the semantic system—Divisible by what? *J. Cognit. Neurosci.* **10:** 766–777.

Petersen, S. E., Fox, P. T., Snyder, A. Z., and Raichle, M. E. 1990. Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. *Science* **249:** 1041–1044.

Price, C. J. 1998. The functional anatomy of word comprehenson and production. *Trends Cognit. Sci.* **2:** 281–288.

Price, C. J. 2000. The functional anatomy of language: Contributions from neuroimaging. *J. Anat.* **197:** 335–359.

Price, C. J., Moore, C. J., Humphreys, G. W., and Wise, R. J. S. 1997. Segregating semantic from phonological processes during reading. *J. Cognit. Neurosci.* **9:** 727–733.

Price, C. J., Wise, R. J. S., and Frackowiak, R. S. J. 1996. Demonstrating the implicit processing of visually presented words and pseudowords. *Cereb. Cortex* **6:** 62–70.

Price, C. J., Wise, R. J., Watson, J. D., Patterson, K., Howard, D., and Frackowiak, R. S. 1994. Brain activity during reading. The effects of exposure duration and task. *Brain* **117:** 1255–1269.

Rugg, M. D., and Henson, R. N. A. 2001. Episodic memory retrieval: An (event-related) functional neuroimaging perspective. In *The Cognitive Neuroscience of Memory Encoding and Retrieval* (A. Parker, E. Wilding, and T. Bussey, Eds.). Psychology Press, London.

Rumsey, J. M., Horwitz, B., Donohue, B. C., Nace, K., Maisog, J. M., and Andreason, P. 1997. Phonological and orthographic components of word recognition. A PET–rCBF study. *Brain* **120:** 739–759.

Schacter, D. L., and Buckner, R. L. 1998. Priming and the brain. *Neuron* **20:** 185–195.

Schacter, D. L., Buckner, R. L., Koutstaal, W., Dale, A. M., and Rosen, B. R. 1997. Late onset of anterior prefrontal activity during true and false recognition: An event-related fMRI study. *NeuroImage* **6:** 259–269.

Seidenberg, M. S., and McClelland, J. L. 1989. A distributed, developmental model of word recognition and naming. *Psychol. Rev.* **96:** 523–568.

Talairach, J., and Tournoux, P. 1988. *Co-planar Stereotaxic Atlas of the Human Brain.* Thieme, Stuttgart.

Thierry, G., Boulanouar, K., Kherif, F., Ranjeva, J. P., and Demonet, J. F. 1999. Temporal sorting of neural components underlying phonological processing. *NeuroReport* **10:** 2599–2603.

Vandenberghe, R., Price, C., Wise, R., Josephs, O., and Frackowiak, R. S. J. 1996. Functional anatomy of a common semantic system for words and pictures. *Nature* **383:** 254–256.

Vazquez, A. L., and Noll, D. C. 1998. Nonlinear aspects of the BOLD response in functional MRI. *NeuroImage* **7:** 108–118.

Wiggs, C. L., and Martin, A. 1998. Properties and mechanisms of perceptual priming. *Curr. Opin. Neurobiol.* **8:** 227–233.