ACADEMIC PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

NeuroImage

NeuroImage 19 (2003) 784–796

www.elsevier.com/locate/ynimg

Within-subject variability of BOLD response dynamics

Jane Neumann,* Gabriele Lohmann, Stefan Zysset, and D. Yves von Cramon

Max-Planck-Institute of Cognitive Neuroscience, Stephanstrasse 1a, D-04103 Leipzig, Germany

Received 6 November 2002; revised 11 February 2003; accepted 3 March 2003

Abstract

In this paper we investigate the within-subject variability of dynamical aspects of the BOLD response obtained in a series of fMRI sessions several days apart. Five different parameters describing the temporal behavior of trial-averaged time courses, such as time-to-peak and time-to-onset, were estimated and analyzed with respect to their variability across nine sessions. Results show that small variances of the estimated parameters can be found, provided that the analysis is restricted to voxels activated in all individual sessions. Among the investigated parameters, time-to-peak shows the most stable behavior. These results were obtained using two different analysis methods, the estimation of the parameters directly from trial-averaged time courses and fitting trial-averaged time courses to an assumed hemodynamic response function. Both methods yield comparable results.

Introduction

Functional magnetic resonance imaging (fMRI) has become an increasingly important method for the study of functional neuroanatomy in humans (Ogawa et al., 1990; Kwong et al., 1992). This noninvasive imaging method is based on changes in the blood oxygenation level which is supposed to be linked to neural activation. However, to date the exact link between increased neural activity and changes in the level of blood oxygenation is not fully understood, making the interpretation of the acquired data difficult. This problem becomes especially apparent when investigating the temporal dynamics of neural activation, with the hemodynamic delay of the blood oxygenation level-dependent (BOLD) response effectively working as a temporal filter on the actual fMRI signal.

When examining fMRI time courses, an increase in the BOLD signal is often observed about 2 s after stimulus onset (DeYoe et al., 1994; Buckner et al., 1996; Buckner, 1998; Menon and Kim, 1999). The BOLD signal reaches its maximum approximately 5 to 8 s after stimulus onset and remains increased beyond the duration of the stimulus (Blamire et al., 1992; DeYoe et al., 1994; Menon et al., 1995). However, a number of studies have shown that this general time course following stimulation can vary considerably when different cortical regions and subjects are compared. Although it is reasonable to assume that the variability of the observed signal reflects at least to some degree the variability of the underlying neural activity, this might not be the only cause. The BOLD response is, for example, sensitive to vessel diameter, whereby longer delays are found for larger vessels (Lee et al., 1995). The temporal behavior of the BOLD signal will thus be partly influenced by differences in the underlying vasculature. Moreover, differences in scanning hardware, procedures, and experimental designs as well as analysis tools and postprocessing strategies will most likely affect any comparative study of brain activity. Knowing the exact amount of these influences on the measured fMRI signal is essential for a further understanding of the neurovascular coupling and the correct interpretation and statistical analysis of the obtained measurements.

The actual source of observable variation in the BOLD signal is manifold, reaching from differences between repeated sessions to variations between cortical regions and subjects. In a number of studies, the temporal behavior of the hemodynamic response has been found stable for repeated trials of a single session (Kim et al., 1997; Aguirre et al., 1998; Miezin et al., 2000). This suggests that averaging

* Corresponding author. Fax: +49-0-341-9940-221.
E-mail address: neumann@cns.mpg.de (J. Neumann).

1053-8119/03/$ – see front matter
doi:10.1016/S1053-8119(03)00177-0

time courses over trials on a voxel-by-voxel basis preserves the temporal properties of the individual trials. However, differences on the order of a few seconds have been observed for estimates of time-to-onset and time-to-peak when comparing trial-averaged time courses between subjects (Kim et al., 1997; Schacter et al., 1997; Buckner et al., 1998; Aguirre et al., 1998; Miezin et al., 2000). Variations of similar proportions have also been found in the timing and shape of the hemodynamic response across different cortical regions of individual subjects. Delays on the order of seconds and prolonged activation were, for example, observed for anterior prefrontal regions relative to visual areas (Schacter et al., 1997; Buckner et al., 1998). Miezin et al. (2000) reported considerable variation between motor and visual cortex of individual subjects, suggesting that the regional variation in the BOLD signal might even be substantially greater than any global factors influencing response properties across subjects. Delays in peak times between different cortical regions, some as long as a few seconds, were further observed by Thierry et al. (1999) and Kruggel and von Cramon (1999b).

Little work has been done addressing the within-subject variability of the BOLD response across a number of different sessions, especially if they are several days or even weeks apart. So far, research into the between-session variability of the BOLD response was largely directed toward the analysis of volume and overlap of activated voxels as well as the magnitude of their activation (Noll et al., 1997; Rombouts et al., 1998; McGonigle et al., 2000; Waldvogel et al., 2000; Maitra et al., 2002). One of the few reports addressing temporal aspects come from Aguirre et al. (1998), who found significant variability in the shape of the hemodynamic response in three of four subjects who performed a simple visually induced motor task during five sessions taking place several days apart.

Knowing the within-subject variability across sessions is essential for the interpretation of results from between-subject analyses. Knowledge of the exact amount of between-session variability is required in order to dissociate the amount of between-subject variability that can truly be attributed to the differences between subjects from the variability simply caused by differences between sessions. Moreover, knowing the amount and sources of within-subject variations across sessions becomes particularly important, for example, for studies addressing learning and habituation effects on a subject's performance as well as aging or recovery of function after brain lesions in patients. Our own work presented in this paper was therefore aimed toward studying the between-session variability for subjects performing the same task in a number of sessions spread out over a few weeks. Given the stable temporal behavior of the BOLD response across trials in a single session, we would expect the within-subject variability between scans to be considerably smaller than the between-subject variability, provided the experimental conditions do not substantially differ from one session to the next. In order to include large parts of the cortex in our analysis, a variant of the Stroop task (Stroop, 1995), known to produce activation in a number of frontal and parietal areas, was chosen as the experimental paradigm.

A second aspect of our investigation addressed the question of whether there exist one or more parameters of the hemodynamic response, i.e., distinct points in the course of the hemodynamic response, that show a particularly consistent temporal behavior. This question was already addressed, for example, by Menon et al. (1998). Using a 4 Tesla scanner, trial-averaged time courses were obtained for V1 after presenting visual stimuli with various delays. An analysis of these time courses showed that only delays in the onset of the time courses correlated with the delay of the stimulus presentation. Other parameters such as time-to-peak and amplitude of the BOLD response exhibited a less stable behavior. This result is in stark contrast with observations reported by Miezin et al. (2000). Comparing the reliability of a number of different parameters of BOLD signals in the visual cortex, they found the estimates of amplitude and time-to-peak to be the most stable ones. This discrepancy clearly needs further investigation.

Finally, we were interested in the influence of the analysis method used to obtain estimates of time lags in the hemodynamic response. Two general approaches can be found in the literature. Specific time points in the BOLD signal, typically time-to-onset and time-to-peak, can be derived directly from preprocessed data as done, for example, by Kim et al. (1997) and Menon et al. (1998) discussed above. Alternatively, parameters of functions assembling the assumed shape of the hemodynamic response and fitted to the acquired data can serve as estimates of such time lags. Examples of this methodology, including linear and nonlinear regressions of Gamma and Gauss functions and linear combinations thereof, can be found in Henson et al. (2002), Liao et al. (2002), Miezin et al. (2000), Kruggel and von Cramon (1999a), and Cohen (1997). These two approaches are subjected to different amounts of numerical instability and interpolation from the measured signal. In order to investigate such influences we applied both methods to our experimental data and compared the results with respect to the reliability of the obtained estimates.

Methods

Subjects and experimental task

For our analysis we used data obtained from an event-related single trial version of the Color–Word Matching Stroop task (Zysset et al., 2001). Four subjects (three female, mean age 23 years) were examined, from whom we obtained written consent prior to the scanning sessions. After instructions and a practice session, nine experimental sessions were performed by each subject within 9 weeks. With one exception, sessions took place on the same day

and time every week. There were three experimental conditions (neutral, congruent, and incongruent). During neutral trials, letters presented in the top row of the screen were 'XXXX' printed in red, green, blue, or yellow, and the bottom row consisted of the color words 'RED,' 'GREEN,' 'BLUE,' and 'YELLOW' printed in black. For congruent trials, the top row consisted of the color words 'RED,' 'GREEN,' 'BLUE,' and 'YELLOW' printed in the congruent color. The incongruent condition was identical to the congruent one, except that the color word was printed in an incongruent color (e.g., 'GREEN' printed in red), in order to produce an interference between color word and color name. The conditions were presented in a randomized order. Stimuli were presented until the subjects responded by tapping the index or the middle finger of the right hand. This led to the presentation of a new stimulus every 6 s on average. A variable onset delay of 0, 400, 800, 1200, or 1600 ms produced an oversampling of the actual image acquisition time of 2000 ms by a factor 5, leading to an acquisition sampling rate of 400 ms.

MRI scanning procedure

The experiment was carried out on a 3T scanner (Medspec 30/100, Bruker, Ettlingen). Sixteen axial slices (19.2 cm FOV, 64 by 64 matrix, 5 mm thickness, 1 mm spacing), parallel to the AC–PC plane and covering the whole brain, were acquired using a single shot, gradient recalled EPI sequence (TR 2000 ms, TE 30 ms, 90° flip angle). One functional run with 648 time points each was conducted, with each time point sampling over the 16 slices. Prior to the functional runs, 16 anatomical T1-weighted MDEFT (Ugurbil et al., 1993; Norris, 2000) images (data matrix 256 × 256, TR 1.3 s, TE 10 ms) and 16 T1-weighted EPI images with the same parameters as the fMRI data were acquired.

fMRI data analysis

The fMRI data were processed with the software LIPSIA (Lohmann et al., 2001). This software package contains tools for preprocessing, registration, statistical evaluation, and presentation of fMRI data.

Functional data were corrected for motion using a matching metric based on linear correlation. To correct for the temporal offset between the slices acquired in one session, a sinc-interpolation based on the Nyquist–Shannon Theorem was applied. A temporal high-pass filter with a cut-off frequency of 1/84 Hz was used for baseline correction of the signal and a spatial Gaussian filter with 4.24 mm FWHM was applied. To align the individual functional data slices onto the corresponding 3D stereotactic coordinate reference system, a rigid linear registration with six degrees of freedom (3 rotational, 3 translational) was performed. The rotational and translational parameters were acquired on the basis of the MDEFT and EPI-T1 slices to achieve an optimal match between these slices and the individual 3D reference data set. This 3D reference data set was acquired for each subject during a previous scanning session. The MDEFT volume data set with 160 slices and 1 mm slice thickness was standardized to the Talairach stereotactic space (Talairach and Tournoux, 1988). The rotational and translational parameters were then used to transform the functional slices using trilinear interpolation, so that the resulting functional slices were aligned with the stereotactic coordinate system.

In order to detect significant activations, a statistical evaluation was performed based on a least-squares estimation using the general linear model for serially autocorrelated observations (Friston, 1994; Worsley and Friston, 1995; Zarahn et al., 1997). The design matrix was generated utilizing a synthetic hemodynamic response function and its first and second derivative (Friston et al., 1998; Josephs et al., 1997). This way, the amplitude as well as temporal aspects are taken into account for the detection of activated voxels. The model equation, including observation data, design matrix, and error term, was convolved with a Gaussian kernel with a dispersion of 4 s FWHM. Contrast maps, i.e., estimates of the raw-score differences between specified conditions and baseline, were generated for each subject and session. For these contrasts one-sample t tests were performed assessing the null hypothesis of zero response, and statistical parametric maps SPM$\{t\}$ were constructed indicating the significance of the response. Obtained t values were subsequently transformed into z values, giving an SPM$\{Z\}$ for each subject and condition. Voxels exceeding the threshold $z = 3.09$, corresponding to $P < 0.001$, were included in the analysis of temporal aspects of the BOLD response. Since the contrast between the neutral and the incongruent conditions represents the main Stroop interference, only these conditions were considered.

In order to further restrict parts of the analysis to voxels showing activation in every individual session, binary masks were obtained for each subject and condition, marking voxels which exceed a threshold of $z = 2.33$, corresponding to $P < 0.01$, in every individual contrast map for a condition and subject. Regions of interest (ROI) were formed within the marked cortical areas, including the local maximum of activation together with all voxels within its 26-adjacency, i.e., voxels whose Euclidean distance from the maximum of activation did not exceed $\sqrt{3} \times$ voxel size. Voxels within this neighborhood that were not activated in all individual sessions were excluded from the ROI.

Analysis of BOLD dynamics

Trial-averaged time courses were obtained on a voxel-by-voxel basis for each session, subject, and condition at a sampling rate of 200 ms, which is twice the rate of image acquisition. Activations for time points falling between two observed points were linearly interpolated from the weighted activation of their neighbors. Four points of interest along the trial-averaged time courses of all subjects,

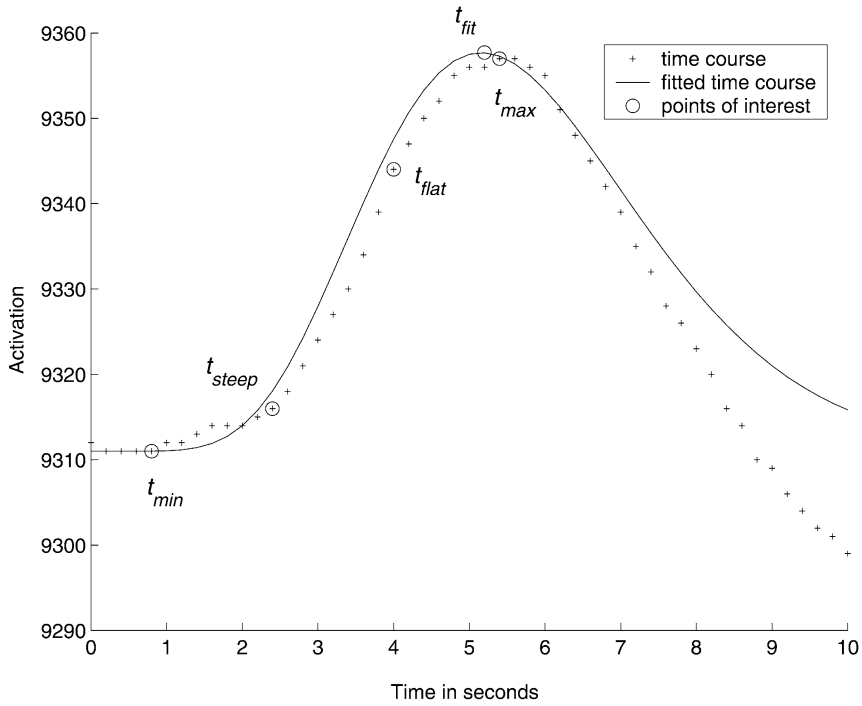

Fig. 1. A typical trial-averaged time course of an activated voxel (Subject 1, Session 2, neutral condition). The original time course and the fitted γ function are shown. The five parameter estimates describing distinct points between stimulus onset and maximum activation in the time course are marked.

scans, and conditions were identified as follows. The minimum and the maximum activation and their respective time lags with reference to stimulation onset, $t_{\min}$ and $t_{\max}$, were sought in the time range of 0 to 5 s and 3 to 8 s, respectively, with $t_{\min} < t_{\max}$. These time ranges were chosen in accordance with earlier reports about the usual time ranges of time-to-onset and time-to-peak (Blamire et al., 1992; DeYoe et al., 1994; Menon et al., 1995; Buckner et al., 1996). In addition, the first and second derivatives of the time courses at each time step were calculated using Taylor polynomial approximations

$$\frac{df}{dx} = \frac{f(x+h) - f(x-h)}{2h} \tag{1}$$

and

$$\frac{d^2f}{dx^2} = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}, \tag{2}$$

with $h = 1$. Time lags of the discrete approximations to the minimum and maximum of the second derivative were then determined by

$$t_{\text{steep}} : \arg\max_x \frac{d^2f}{dx^2} \tag{3}$$

and

$$t_{\text{flat}} : \arg\min_x \frac{d^2f}{dx^2} \tag{4}$$

for all $x \in [0 \ldots n]$, whereby

$$\frac{df}{dx}(x_{t_{\text{steep}}}) > 0, \quad \frac{df}{dx}(x_{t_{\text{flat}}}) > 0 \quad \text{and} \quad t_{\text{steep}} < t_{\text{flat}}. \tag{5}$$

t_{steep} marks the point along the time course where the BOLD response starts rising steeply from the baseline after the presentation of a stimulus. This point can be interpreted as onset of the response function. t_{flat} marks the point along a time course where the function flattens out again before reaching the maximum of the activation. This point should be in close proximity to time-to-peak, but unlike time-to-peak should not be prone to estimation errors caused by prolonged activations. Such prolonged activations result in plateaus in the estimated BOLD response which make the exact identification of time-to-peak difficult. However, the length of activation should not affect t_{flat}, as this always marks the beginning of a period of increased activation, be it a single well-defined peak or a plateau of several seconds.

A typical trial-averaged time course with the four points of interest marked is shown in Fig. 1. Note that reliably estimating t_{steep} and t_{flat} required smoothing the time courses using a Gauss filter with $\sigma = 1$, thereby reducing the number of local extrema of the first and second derivatives.

For each subject and condition, the values for the four points of interest were averaged across the nine sessions on a voxel-by-voxel basis. Mean $\hat{t}$ and standard deviation σ for each point were color-coded and represented in intersection

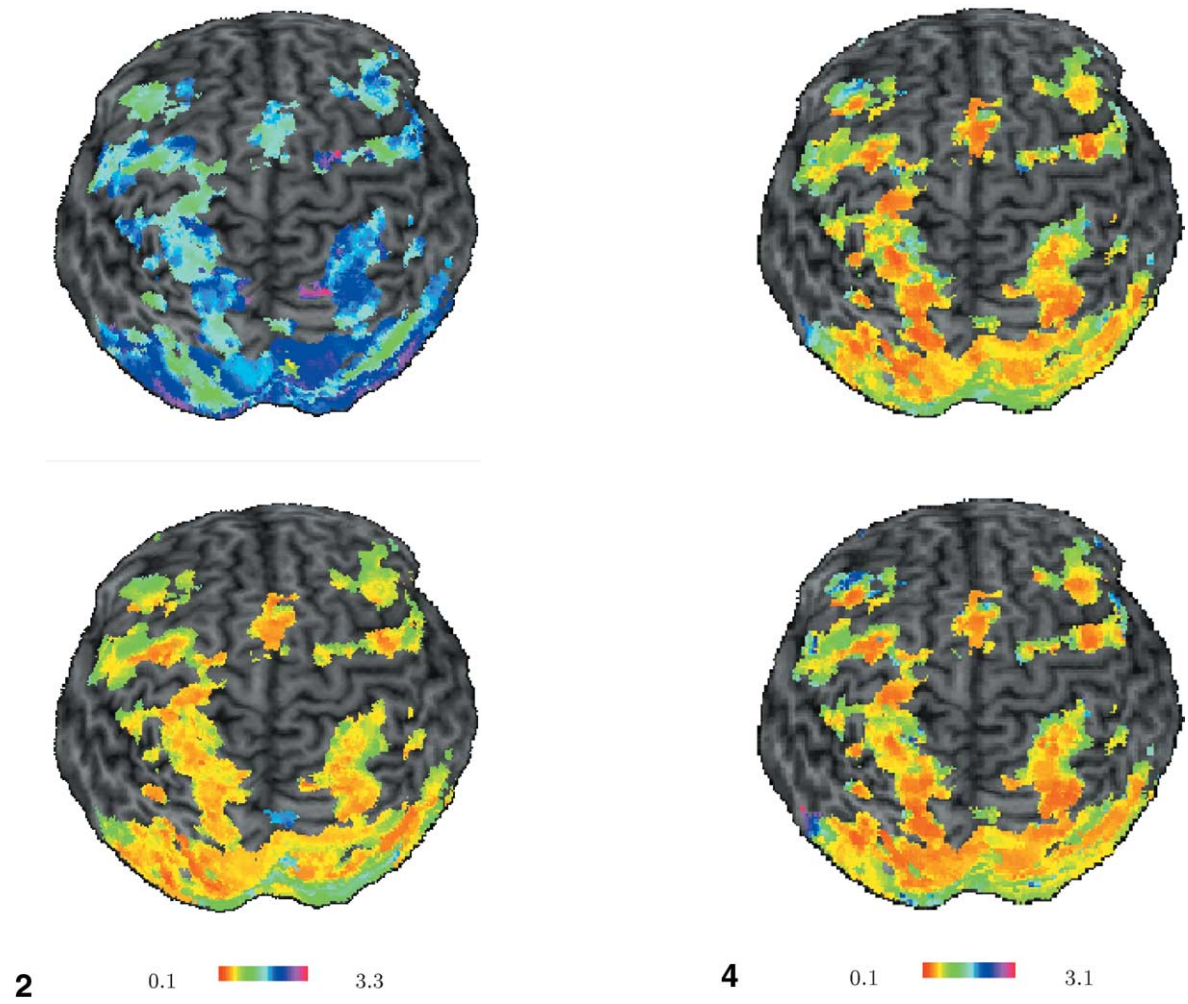

Fig. 2. Intersection flat maps showing means $\hat{t}_{\text{steep}}$ (top) and standard deviations σ_{steep} (bottom) for the incongruent condition in Subject 4 after averaging across all sessions. Only voxels exceeding a z score of 3.09 in the corresponding SPM{Z} are shown. Mean values vary in a range of 0.2 to 3.3 s in different cortical regions. Standard deviations obtained across all sessions range from 0.1 to 2.33 s, but exceed 1 s in some cases only.
Fig. 4. The two estimates for the variability of time-to-peak, σ_{max} and σ_{fit}, shown for the incongruent condition in Subject 4. Visual inspection already reveals that the two estimates obtained by two different methods provide nearly identical values.

flat maps (Lohmann et al., 2002). These maps describe a projection of cortical regions onto a 2D plane which minimizes geometrical distortion along the lateral left–right direction and allows for a convenient inspection of wide parts of both hemispheres.

Our second method of analysis was chosen based on the observation that specific parameters or time points of the hemodynamic response can be derived directly from parameters of a model function fitted to the acquired data. It is believed that the course of the BOLD signal can be reasonably well approximated by a γ function (Friston et al., 1994; Boynton et al., 1996; Lange and Zeger, 1997). Friston et al. (1998) suggested a sum of two γ functions to account for the often observed undershoot following maximum activation. Since we were primarily interested in the temporal behavior of the hemodynamic response up to the maximum activation, we used only a single γ function. However, in preliminary studies we found fitting unsatisfactory when using the usually applied three-parameter γ function (Cohen, 1997; Miezin et al., 2000). For about 10% of all activated voxels, no fit could be obtained within 10,000 iterations of the fitting procedure. Changes in parameter initialization had only marginal impact on these results. Moreover, visual inspection of the fitted functions revealed that, although in many cases the obtained fit resembled the shape of the underlying time course, the amplitude in particular was often not very well approximated. We therefore introduced an additional parameter which allowed a more flexible modulation of the amplitude of the function and fitted the four-parameter function

$$f(x) = \left(\frac{x}{a}\right)^{b} e^{-(x-c)/d} \tag{6}$$

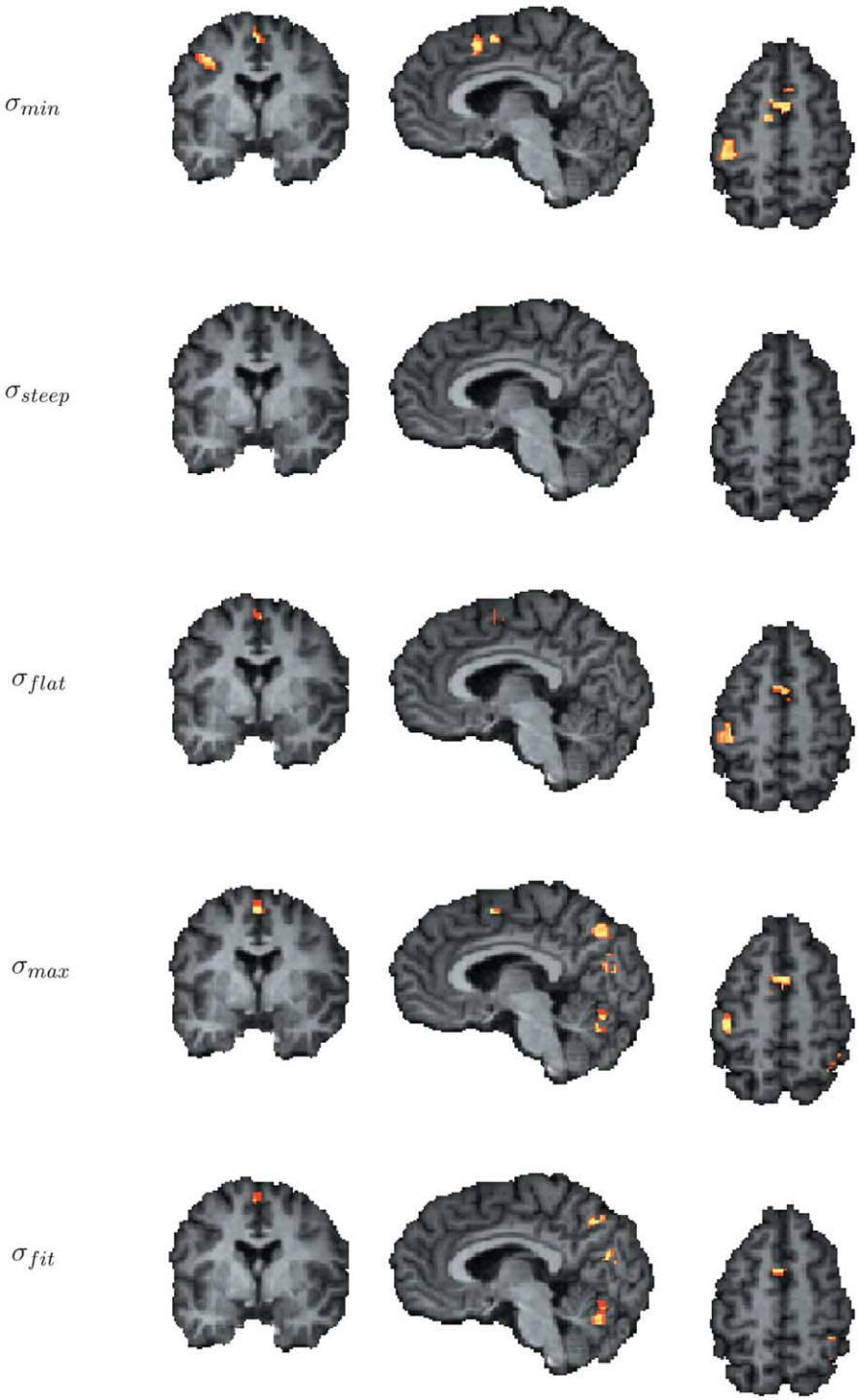

Fig. 3. Cortical regions with standard deviations $\sigma < 0.5$ found for the five estimates in the neutral condition in Subject 1. Regions with small variance largely overlapped with highly activated areas but differed between the five points along the BOLD response. In this example σ_{steep} produced no such region.

leaving all four parameters subject to the optimization. Data were fitted to the model and the model parameters estimated using the Levenberg–Marqhardt algorithm, a standard routine for nonlinear least-squares minimization. For discussion of the algorithm see, e.g., Seber and Wild (1989). The standard deviations obtained for all data points when calculating the trial-averaged time courses served as additional input to the fitting procedure. In this way, the adaptable parameters were forced to fit more reliably measured points better than highly unstable ones. In order to ensure that data were fitted best in the area of the four points of interest, the time range for fitting was restricted from the minimum of the average time course to 1 s after the maximum activation. From Eq. (6) the product of the parameters b and d can be

Table 1
Relative number of voxels where each estimate showed the smallest standard deviation

Subject	Condition	Voxels	σ_{min}	σ_{steep}	σ_{flat}	σ_{max}	σ_{fit}
1	Incongruent	11689	22.59%	6.60%	15.39%	23.37%	32.05%
	Neutral	6769	21.23%	2.98%	16.65%	29.71%	29.43%
2	Incongruent	12658	14.12%	4.09%	20.87%	22.67%	38.25%
	Neutral	10341	11.81%	2.82%	32.88%	20.10%	32.38%
3	Incongruent	15104	10.74%	2.93%	10.18%	20.74%	55.41%
	Neutral	14081	10.18%	2.25%	10.87%	26.59%	50.10%
4	Incongruent	8284	28.15%	7.50%	13.55%	18.60%	32.20%
	Neutral	6982	25.78%	6.86%	18.17%	22.16%	27.03%

Note. A comparison of the standard deviations of all five estimates for the incongruent and neutral conditions in all subjects. Each row shows the number of activated voxels and the relative number of voxels where each of the five estimates showed the smallest standard deviation (summing up to 100%). In most cases, time-to-peak estimated from the fitted γ function shows the highest number, i.e., is the most stable among the five estimates. For the two exceptions, the neutral condition in Subject 1 and Subject 2, it is the second stable estimate with nearly the same share of voxels as the first.

taken directly as an estimate for the time-to-peak of the fitted data (Cohen, 1997; Glover, 1999; Liao et al., 2002). This point, subsequently called t_{fit}, was chosen for comparison with the results of our first analysis method. It is shown together with the four direct estimates in Fig. 1.

The least variable of the five obtained estimates was found by comparing their standard deviations after averaging across all sessions of a subject on a voxel-by-voxel basis. For each estimate the number of voxels for which it showed the smallest standard deviation was counted.

Results

A typical distribution of the mean time lags and respective standard deviations of voxels exceeding a z score of 3.09 in the corresponding SPM$\{Z\}$ is exemplified in Fig. 2 for time-to-onset obtained for the incongruent condition in Subject 4. Representing the results in an intersection flat map provides an overview of the time lags over the whole brain, also showing the variability of the estimates between different cortical regions. Mean time lags $\hat{t}_{steep}$ in this example range from 0.20 to 3.30 s with σ_{steep} between 0.14 and 2.33 s. Similar time differences in mean values between voxels were observed for the other subjects and conditions. Roughly speaking $\hat{t}_{min}$ and $\hat{t}_{steep}$ were found to be between 0 and 4 s and between 0.1 and 4.4 s, respectively. Values for $\hat{t}_{flat}$ and $\hat{t}_{max}$ lay between 3 and 7.6 s and between 3.3 and 7.9 s, respectively. The values of the obtained standard deviations were usually in a range of 0.1 to 2 s, in most cases below 1 s, but in single cases as high as 4 s. A systematic difference in the estimates between the examined conditions could not be observed.

Note that the individual values t_{min}, t_{steep}, t_{flat}, and t_{max} obtained before averaging across sessions were distributed over the entire time range permitted, i.e., between 0 and 5 s for t_{min} and t_{steep} and between 3 and 8 s for t_{flat} and t_{max}, respectively. The values of t_{fit} were not restricted and even exceeded these boundaries. These widespread values can be explained by the fact that some voxels included in the SPM$\{Z\}$ with a relatively small value were not activated in all individual sessions. For nonactivated voxels, however, the shape of the time course might not be well approximated by the γ function and the points of interest can be placed outside the assumed range. This in turn accounts for the relatively high standard deviations observed in some cases when averaging across sessions.

The large variations observed for at least some voxels lead us to believe that the within-subject variability of the BOLD response is not significantly smaller than the between-subject variability observed in the literature. However, a closer inspection of the obtained estimates reveals that some cortical areas containing voxels with consistently small variation can be identified in all four subjects. This is exemplified in Fig. 3 for the neutral condition of Subject 1. Areas containing voxels with standard deviation $\sigma < 0.5$ are shown for all estimates. This particular threshold was chosen to be notably smaller than the variance of response estimates across subjects, which is typically as large as a few seconds.

With the exception of $\hat{t}_{steep}$, all estimates produced a number of cortical areas with such small temporal variation in the BOLD response. Although not completely identical, these areas largely overlapped with highly activated areas. They include the presupplementary motor area, the left inferior frontal sulcus, the left intraparietal sulcus, and the left inferior temporal gyrus, which have also produced distinguished activations in previous Stroop studies (Zysset et al., 2001). As can be seen, however, different points of interest produced different such areas, indicating that the variances in the BOLD response do not only depend on the cortical region but might also vary along the time course of the signal.

Since we were interested in the question of which of the five parameters can be most reliably estimated and shows the most stable temporal behavior, a comparison of the standard deviations of all five estimates was performed for all subjects. The results are presented in Table 1. For each subject and condition the number of activated voxels was counted and the relative number of voxels (in percentage)

where an estimated parameter produced the smallest standard deviation was determined. For example, 11,689 voxels were activated for the incongruent condition in Subject 1. For 32.05% of these voxels, σ_{fit} was smaller than the standard deviations of the other four estimates, which makes $\hat{t}_{\text{fit}}$ the least variable point along the trial-averaged time courses of this subject. $\sigma_{\max}$ was the smallest of all five estimates for 23.37% of all voxels, followed by $\sigma_{\min}$ with 22.59% of all voxels, and so on. As can be seen in Table 1, with two exceptions σ_{fit} was most often found to be the smallest of the five estimated standard deviations usually with about 30%, in Subject 3 with even over 50% of all activated voxels. This means that $\hat{t}_{\text{fit}}$, i.e., time-to-peak estimated from the fitted γ function, was the most stable point along the average time courses. Among the four estimates obtained directly from the time courses, there also seems a tendency for time-to-peak to be the least variable point. This agrees with the observations reported by Miezin et al. (2000) who found time-to-peak the most stable point along trial-averaged time courses. Our data do not support the observations by Menon et al. (1998), who reported the early part of the hemodynamic response to show stable dynamic behavior. On the contrary, $\hat{t}_{\text{steep}}$ in particular was by far the least reliable estimate for all subjects. We note, however, that Menon's analysis is not directly comparable to our results as the time-to-onset was determined by different methods in the two studies.

For the comparison of the two analysis methods, obtaining estimates directly from the averaged time courses and by means of fitting the data to a γ function, visual inspection of intersection flat maps for $\sigma_{\max}$ and σ_{fit} proved sufficient. This is exemplified in Fig. 4 for the incongruent condition in Subject 4. As becomes immediately obvious, standard deviations of both estimates are of the same magnitude. Similar results were obtained for all other subjects and conditions. We would thus argue that neither of the two approaches is generally to be preferred over the other when investigating the variability of parameters of the BOLD response describing its dynamic aspects.

As our earlier results show, using statistical parametric maps to mask activated voxels for further analysis is problematic, because they might include voxels that do not show activations in all individual sessions. We would expect considerably less variation in the parameters describing the BOLD dynamics when only voxels which are activated in all sessions are considered. We thus restricted another step of our analysis to cortical regions only containing such voxels. We chose the incongruent condition for this analysis, which produced the strongest activations in all four subjects. Binary maps were obtained for all subjects marking voxels that exceeded the threshold $z = 2.33$ in all contrast maps of the incongruent condition. Five cortical areas where found to contain such voxels in all four subjects. These areas are shown in Fig. 5 and include the presupplementary motor area (preSMA), the left and right intraparietal sulcus (L SIP and R SIP), and the left and right inferior precentral sulcus (L IPCS and R IPCS). ROIs were formed within these areas as described in the previous section, and average time courses for these regions, the estimates of the points of interest along the time courses, and their respective means and standard deviations across the nine sessions were obtained. The standard deviations for all estimates in the five ROIs of each subject are listed in Table 2.

As expected, for these consistently activated voxels the standard deviations of the time lags are considerably below 1 s, often even below 0.5 s. An exception is σ_{steep}, the standard deviation of time-to-onset. However, time-to-onset was already found to be by far the least stable point along the BOLD response, shown by the small values for σ_{steep} in Table 1. Overall, the small standard deviations show that estimates of parameters describing the temporal behavior of the BOLD response can be meaningfully averaged across a number of sessions of the same subject as long as this averaging is performed only on voxels activated in all individual sessions.

Note that although the small number of subjects does not provide the means for a between-subject analysis, the obtained standard deviations were at least of the same magnitude when compared between subjects. However, the mean values of the estimates, shown in Table 3, varied considerably across cortical regions and across subjects. This is true even for the most stable estimate $\hat{t}_{\text{fit}}$, as can be seen in Fig. 6. Subjects 2 and 4 show a faster BOLD response in all cortical regions than Subjects 1 and 3. Moreover, mean time-to-peak varies considerably between cortical regions in all four subjects and, most notably, the order of activation in the five cortical regions differs immensely between subjects. This supports earlier observations by Miezin et al. (2000), who found that absolute estimates of time-to-peak and time-to-onset in the hemodynamic response have only a rough relation to the likely ordering of neural activity in different cortical regions. Despite the relatively low variance of these estimates within the same subject, their exact interpretation thus still remains an open question.

Discussion

The main findings of this experiment can be summarized in three points. First, the between-session variability of the hemodynamic response is comparatively small, usually well below 1 s, for voxels activated in every individual session. The between-session variability of the BOLD response for a single subject is thus much smaller than the variability between subjects, where estimates of time-to-onset and time-to-peak are usually reported to vary in a range of a few seconds. Second, among different points along a trial-averaged time course, describing its temporal behavior from stimulus onset to maximum activation, time-to-peak shows at least a tendency to be the least variable one. Finally, the two common approaches to estimating parameters of the

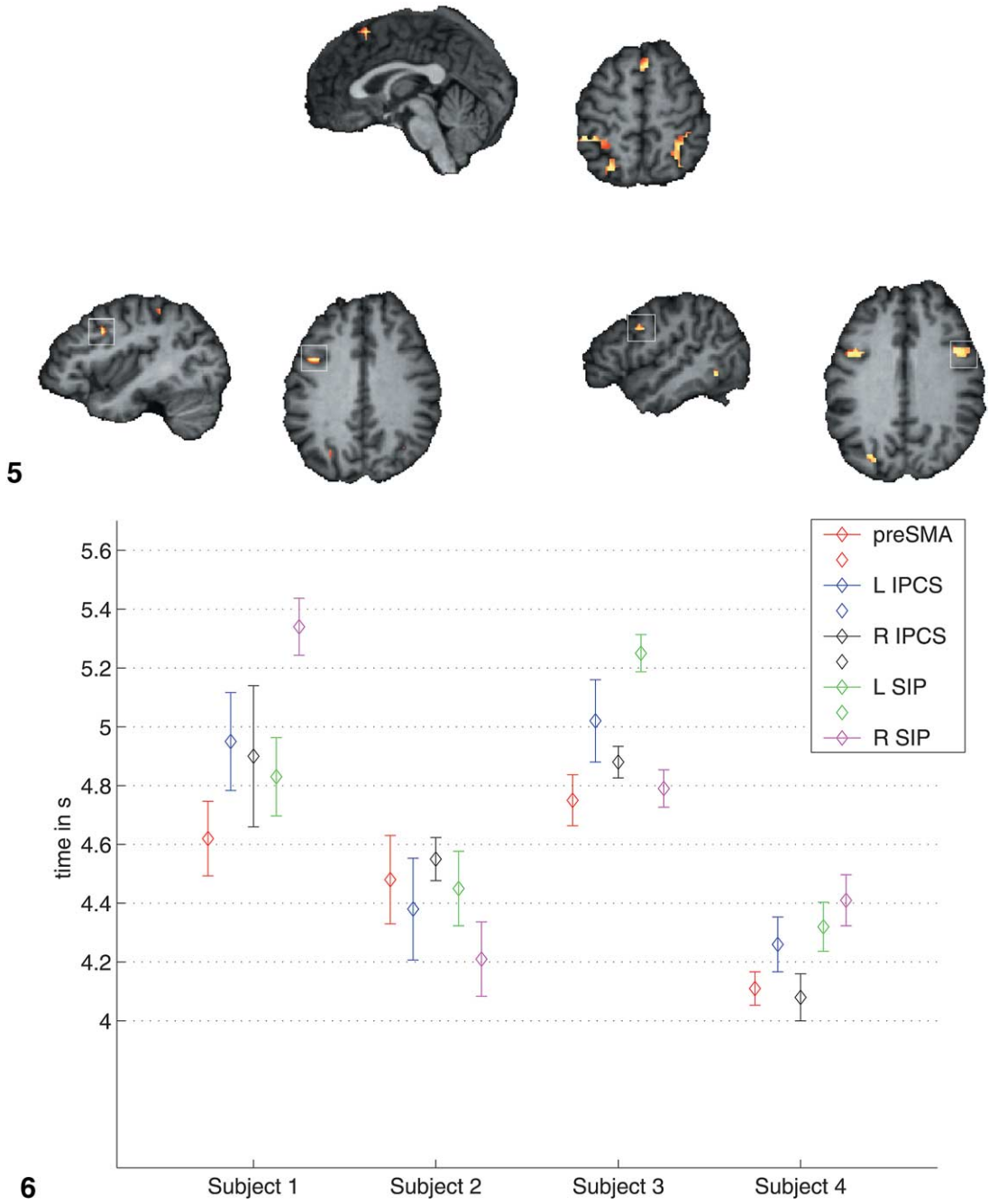

Fig. 5. Five cortical regions were found activated in all individual sessions for the incongruent condition in all four subjects. The regions are shown here for Subject 4 and include presupplementary motor area, left and right intraparietal sulcus (top row), and left and right inferior precentral sulcus (bottom row left and right, respectively).

Fig. 6. Mean values and standard errors of the time lags for the center of activation in the 5 ROIs of all subjects. The most stable estimate $\hat{t}_{\text{fit}}$ obtained for the incongruent condition is shown. Even for this estimate, mean values vary significantly between cortical regions in single subjects and even more so between different subjects. Also note a different temporal order of the five cortical regions in all subjects.

BOLD response, deriving them directly from trial-averaged time courses or from functions fitted to the acquired data, yield nearly identical results when applied to our data. Thus, neither of the two methods is generally to be preferred over the other.

The lack of reports concerning the within-subject variability of the BOLD response across different scanning sessions can probably be attributed to the relatively high effort, time, and expense required for such long-term studies. Moreover, as for comparisons across subjects, the analysis of temporal aspects of the hemodynamic response is hampered by the relatively low reproducibility of activations and a high variability in activation magnitude across different sessions. Noll et al. (1997), for example, found

Table 2
Standard deviations of all estimates in five cortical regions

Subject	Estimate	preSMA	L IPCS	R IPCS	L SIP	R SIP
1	σ_{min}	0.33	0.36	0.59	0.54	0.52
	σ_{steep}	0.69	0.74	0.55	0.97	1.15
	σ_{flat}	0.57	0.70	0.63	0.30	0.65
	σ_{max}	0.52	0.60	0.55	0.33	0.36
	σ_{fit}	0.40	0.56	0.74	0.30	0.34
2	σ_{min}	0.45	0.34	0.61	0.28	0.54
	σ_{steep}	0.87	0.72	0.99	0.68	1.17
	σ_{flat}	0.27	0.14	0.28	0.59	0.37
	σ_{max}	0.57	0.40	0.33	0.61	0.48
	σ_{fit}	0.47	0.49	0.26	0.41	0.41
3	σ_{min}	0.35	0.69	0.55	0.62	0.55
	σ_{steep}	0.79	0.94	1.07	0.44	0.84
	σ_{flat}	0.35	0.48	0.50	0.68	0.26
	σ_{max}	0.24	0.26	0.18	0.37	0.15
	σ_{fit}	0.20	0.38	0.25	0.19	0.19
4	σ_{min}	0.18	0.22	0.39	0.26	0.37
	σ_{steep}	0.21	0.48	0.83	0.55	0.65
	σ_{flat}	0.19	0.24	0.31	0.13	0.26
	σ_{max}	0.19	0.28	0.27	0.22	0.25
	σ_{fit}	0.18	0.33	0.33	0.27	0.23

Note. Standard deviations (in s) of the five estimates in five comparable cortical regions for all subjects. Standard deviations for most estimates are considerably below 1 s, often found between 0.1 and 0.5 s. Only time-to-onset (σ_{steep}) shows higher variation for all four subjects.

considerably less reliability of voxel activation between sessions than between scans in a single session. Rombouts et al. (1998) reported a ratio of overlapping activated voxels as low as 64% for scans of the same subject obtained in sessions several days apart. Waldvogel et al. (2000) observed a variability of the number of activated voxels across sessions as high as 1150% and of the activation amplitude of 250%. Most notably, in an extensive long-term study McGonigle et al. (2000) acquired data from 99 sessions of a single subject presented with simple visual, motor, and cognitive paradigms. For all tasks the pattern of activation varied widely between repeated sessions, with some sessions showing no significant activation at all. Thus, even a large number of repeated sessions makes inferences about general patterns of activation difficult. High variability of the magnitude and pattern of activation must clearly influence the analysis of temporal aspects of the hemodynamic response, however.

We would argue that the relatively high variances we initially observed for some voxels was largely caused by the fact that these voxels were not consistently activated in all sessions. As can be seen in Fig. 4 for σ_{max} and σ_{fit}, such voxels were usually located close to the edge of clusters of activated voxels. Time courses of such voxels often did not possess the parameters we wished to analyze, however. Frequently, we found a number of local minima before a larger signal increase, and the signal did not rise steadily, causing a large number of local extrema in the second derivatives. In some cases we observed two or even three peaks of activation, and some time courses did not show any increase of activation in the assumed time range at all. This of course makes providing sensible estimates for parameters like time-to-peak and time-to-onset difficult. Moreover, the γ function fits only poorly to such data. Thus, given the relatively small number of sessions investigated, estimates from a voxel with no activation in only one session can already increase the variance of the averaged data considerably. The two methods of analysis, though they can theoretically be applied to the whole brain, therefore only yield interpretable results for voxels activated in all investigated sessions.

It has been argued that model-based analysis methods for functional magnetic resonance images are hampered by the fact that they make specific assumptions or require a priori knowledge about the shape of the time courses to be investigated (Duann et al., 2002; McKeown et al., 1998). Building upon such model-based approaches in turn restricts any further analysis of temporal aspects of the BOLD response to voxels whose time course correlates well with a predicted response function. Recently, Duann et al. (2002) demonstrated that when abandoning any a priori assumptions about the shape of the hemodynamic response and adopting data-driven analysis approaches like ICA, one finds marked variations of the derived components not only between subjects but also across stimulus types, sessions, and within sessions across trials, suggesting even higher variability of the hemodynamic response than observed with model-based analysis methods. However, components derived by PCA or

Table 3
Mean values of all estimates in five cortical regions

Subject	Estimate	preSMA	L IPCS	R IPCS	L SIP	R SIP
1	$\hat{t}_{min}$	0.33	0.33	0.40	0.62	0.73
	$\hat{t}_{steep}$	1.16	1.42	1.29	1.24	1.78
	$\hat{t}_{flat}$	4.20	4.33	4.62	4.38	5.04
	$\hat{t}_{max}$	4.71	4.91	4.87	4.89	5.47
	$\hat{t}_{fit}$	4.62	4.95	4.90	4.83	5.34
2	$\hat{t}_{min}$	0.49	0.16	0.31	0.11	0.27
	$\hat{t}_{steep}$	0.96	0.40	0.82	0.29	1.38
	$\hat{t}_{flat}$	4.16	4.07	4.13	4.00	4.04
	$\hat{t}_{max}$	4.53	4.42	4.69	4.51	4.27
	$\hat{t}_{fit}$	4.48	4.38	4.55	4.45	4.21
3	$\hat{t}_{min}$	0.73	0.73	0.33	0.56	0.40
	$\hat{t}_{steep}$	1.22	1.29	1.07	2.16	1.20
	$\hat{t}_{flat}$	4.40	4.44	4.31	4.36	4.18
	$\hat{t}_{max}$	4.73	5.04	4.96	5.20	4.73
	$\hat{t}_{fit}$	4.75	5.02	4.88	5.25	4.79
4	$\hat{t}_{min}$	0.40	0.24	0.24	0.18	0.31
	$\hat{t}_{steep}$	0.89	0.76	0.80	0.96	0.67
	$\hat{t}_{flat}$	4.00	4.20	4.07	4.00	4.07
	$\hat{t}_{max}$	4.27	4.38	4.31	4.29	4.40
	$\hat{t}_{fit}$	4.11	4.26	4.08	4.32	4.41

Note. Mean values (in s) of the five estimates in five comparable cortical regions for all subjects. Values vary considerably across cortical regions and subjects. A comparison of the mean values of the most stable estimate $\hat{t}_{fit}$ can also be seen in Fig. 6.

ICA usually lack a clear physical interpretation. This makes a direct comparison of their variability to the variability of parameter estimates of the BOLD response like time-to-peak and time-to-onset impossible. It should also be noted that while abandoning a priori assumptions about the shape of the hemodynamic response, data-driven analysis methods are still based on assumptions which are not guaranteed to be met by fMRI data sets (Stone et al., 2002). PCA and ICA, for example, assume orthogonality and spatial independence of the derived components, respectively. Results from these methods thus depend crucially on the validity of such assumptions, just as the success of a model-based analysis depends on the correctness of the assumed hemodynamic response model. We still regard model-based approaches as appropriate tools for the analysis of BOLD response dynamics. This view is further supported by the point raised above, namely that only restricting the analysis to activated voxels, i.e., to voxels with time courses that roughly follow the assumed shape of a model function of the hemodynamic response, provides interpretable and comparable estimates of parameters describing the temporal behavior of the BOLD response.

The actual method of finding parameters (amplitude, time-to-peak, time-to-onset, etc.) that make comparisons of BOLD responses possible can also be viewed in the light of model-based versus data-driven methods. Such parameters or time points can be derived directly from preprocessed and (sometimes) averaged data as done with our first method and also, for example, in the work by Kim et al. (1997) and Menon et al. (1998) discussed at the beginning of the paper. Alternatively, parameters of functions fitted to the acquired data can serve as estimates for parameters of the BOLD response. This was demonstrated, for example, by Liao et al. (2002), Henson et al. (2002), Kruggel and von Cramon (1999a), and Miezin et al. (2000), and in our second analysis method. For our data, both methods yield comparable results, leaving the choice of method to the experimenter. It is worth pointing out, though, that both methods have their advantages and disadvantages. As argued above, the former approach is difficult to implement, if the data differ widely from the assumed shape and time range of the hemodynamic response. If these underlying assumptions are met, however, the derived parameters should be very accurate and close to reality, as the method operates directly on the acquired data, with few approximations and interpolations from the measured signal. In contrast, the method of fitting a function to the acquired data provides estimates of temporal properties of the BOLD response depending on parameters which are in turn estimated, and might thus be prone to high estimation errors. This was explicitly pointed out by Henson et al. (2002), who described latency differences across trial types by the ratio of two parameters of a fitted canonical function and its derivative. On the other hand, fitting a function to the data has the advantage that obtained parameters of the model functions can directly (and quickly) serve as descriptions of temporal aspects of the underlying data, with good fitting procedures provided. This is particularly true for fitting nonlinear functions like the γ function to the data, where some of the obtained parameters have a clear physical interpretation such as time-to-peak or amplitude of the activation. This immediate interpretation of the fitted parameters is more difficult when fits of linear combinations of functions are employed.

Somewhat surprisingly, we observed a higher variability for the estimated time-to-onset than for all other points along the trial-averaged time courses. Moreover, the variability of $\hat{t}_{\text{flat}}$ was not, as initially expected, generally smaller than that of time-to-peak. We would argue that these results are caused by the fact that it is relatively difficult to exactly determine these two points from real fMRI signals. While it is possible to unambiguously determine $\hat{t}_{\text{steep}}$ and $\hat{t}_{\text{flat}}$ from a model function like γ, the method becomes less accurate when applied to real data, as the first and second derivatives of the estimated hemodynamic responses still contain a large number of local extrema. The high variability of the two points might thus not so much reflect their instability in the measured signal but the still insufficient means used for their exact determination. Methods for a more accurate identification of the two points from the obtained measurements will have to be the subject of further research.

While we could present results showing that trial-averaged time courses have a relatively stable temporal behavior within a single subject, the comparison of time courses obtained from different subjects remains problematic. In their early work Buckner et al. (1998) and Schacter et al. (1997) observed that the estimated hemodynamic response functions were relatively stable when compared across independent groups of subjects. Time courses averaged over groups of six and seven subjects, respectively, nearly overlapped. These observations were recently supported by Miezin et al. (2000), who reported some strong tendencies present in the hemodynamic responses of different subjects which deem averaging across a relatively large number of subjects reasonable. However, such tendencies could not be found in our four subjects. Mean response times and the response order of five cortical regions differed considerably from subject to subject. While the number of subjects in this experiment is too small for any statistical analysis, these results still demand particular caution for the interpretation of hemodynamic response functions averaged across subjects or different cortical regions.

Note finally that the correctness of a comparison of the BOLD response across different cortical regions and subjects will depend on the parameters or entities used for the comparative studies. We would argue that a parameter which can be reliably estimated for a region in a single subject is the natural choice for the detection of differences between regions and subjects. Our results suggest that time-to-peak is such a parameter. However, one has to keep in mind that the reasons for the stability of some parameters and the instability of others are still not fully understood.

Stability of a parameter could, for example, be caused by saturation effects influencing the neurovascular coupling. Although such effects are more conceivable for the amplitude of an activation than for temporal aspects, they so far cannot be excluded. While in the absence of better knowledge a stable parameter should be the best choice for comparative studies, the actual physical interpretation of the stability still needs further research.

Conclusion

We have presented an experiment addressing the within-subject variability of temporal aspects of the hemodynamic response across a number of sessions several days and weeks apart. For four subjects time lags of four points along the time course of activated voxels were analyzed using two different methods. The results of both methods applied to the whole brain revealed relatively high variance across scanning sessions for at least some voxels. Estimates for different points along trial-averaged time courses varied for some voxels in a range of a few seconds, which is comparable to the between-subject variability of the hemodynamic response reported in the literature. However, these relatively unstable voxels, although exceeding the commonly used threshold in the SPM $\{Z\}$ obtained from all sessions, were not found activated in all individual sessions. For voxels activated in all individual sessions the between-session variability of the BOLD response was much smaller. Variances as small as 0.1 s were found, suggesting a constant temporal behavior of the BOLD signal in cortical regions with stable activation across all sessions. The complete physiological and functional interpretation of the observed variability of the hemodynamic response and the exact relation of between-subject variability of the BOLD signal to the between-session variability of single subjects still remain open research questions, however, and will be the subject of our future work.

References

Aguirre, G.K., Zarahn, E., D'Esposito, M., 1998. The variability of human BOLD hemodynamic responses. NeuroImage 8 (4), 360–369.

Blamire, A., Ogawa, S., Ugurbil, K., Rothman, D., McCarthy, G., Ellermann, J., Hyder, F., Rattner, Z., Shulman, R., 1992. Dynamic mapping of the human visual cortex by high-speed magnetic resonance imaging. Proc. Natl. Acad. Sci. USA 89, 11069–11073.

Boynton, G., Engel, S., Glover, G., Heeger, D., 1996. Linear systems analysis of fMRI in human VI. J. Neurosci. 16, 4207–4221.

Buckner, R., Bandettini, P., O'Craven, K., Savoy, R., Petersen, S., Raichle, M., Rosen, B., 1996. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. Proc. Natl. Acad. Sci. USA 93, 14878–14483.

Buckner, R., Koutstaal, W., Schacter, D., Dale, A., Rotte, M., Rosen, B., 1998. Functional-anatomic study of episodic retrieval. II. Selective averaging of event-related fMRI trials to test the retrieval success hypothesis. NeuroImage 7, 163–175.

Buckner, R.L., 1998. Event-related fMRI and the hemodynamic response. Hum. Brain Mapp. 6, 373–377.

Cohen, M.S., 1997. Parametric analysis of fMRI data using linear systems methods. NeuroImage 6 (2), 93–103.

DeYoe, E., Bandettini, P., Neitz, J., Miller, D., Winans, P., 1994. Functional magnetic resonance imaging (fMRI) of the human brain. J. Neurosci. Methods 54 (2), 171–187.

Duann, J.-R., Jung, T.-P., Kuo, W.-J., Yeh, T.-C., Makeig, S., Hsieh, J.-C., Sejnowski, T., 2002. Single-trial variability in event-related BOLD signals. NeuroImage 15, 823–835, doi:10.1006/nimg.2001.1049.

Friston, K., 1994. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.

Friston, K., Fletcher, P., Joseph, O., Holmes, A., Rugg, M., Turner, R., 1998. Event-related responses in fMRI: characterising differential responses. NeuroImage 7, 30–40.

Friston, K., Jezzard, P., Turner, R., 1994. Analysis of functional MRI time series. Hum. Brain Mapp. 1, 153–171.

Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9, 416–429 doi:10.1006/nimg.1998.419.

Henson, R., Price, C., Rugg, M., Turner, R., Friston, K., 2002. Detecting latency differences in event-related BOLD responses: application to words versus nonwords, and initial versus repeated face presentations. NeuroImage 15, 83–97, doi:10.1006/nimg.2001.0940.

Josephs, O., Turner, R., Friston, K., 1997. Event-related fMRI. Hum. Brain Mapp. 5, 243–248.

Kim, S.-G., Richter, W., Ugurbil, K., 1997. Limitations of temporal resolution in functional MRI. Magn. Reson. Med. 37 (4), 631–636.

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

Kruggel, F., von Cramon, D., 1999b. Temporal properties of the hemodynamic response in functional MRI. Hum. Brain Mapp. 8, 259–271.

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

Lange, N., Zeger, S., 1997. Non-linear fourier time series analysis for human brain mapping by functional magnetic resonance imaging. Appl. Stat. 46, 1–29.

Lee, A., Glover, G., Meyer, C., 1995. Discrimination of large venous vessels in time-course spiral blood-oxygen-level-dependent magnetic-resonance functional neuroimaging. Magn. Reson. Med. 33, 745–754.

Liao, C., Worsley, K., Poline, J.-B., Aston, J., Duncan, G., Evans, A., 2002. Estimating the delay of the response in fMRI data. NeuroImage 16, 593–606, doi:10.1006/nimg.2002.1096.

Lohmann, G., Mueller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., Zysset, S., von Cramon, D., 2001. Lipsia—a new software system for the evaluation of functional magnetic resonance images of the human brain. Comput. Med. Imag. Graphics 25 (6), 4498–4457.

Lohmann, G., Schubotz, R., von Cramon, D.Y., 2002. Visualization of fMRI data using coronal flat maps. [Abstract] Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan. Available on CD-Rom in NeuroImage, Vol. 16, No. 2.

Maitra, R., Roys, S., Gullapalli, R., 2002. Test-retest reliability estimation of functional MRI data. Magn. Reson. Med. 48, 62–70.

McGonigle, D., Howseman, A., Athwal, B., Friston, K., Frackowiak, R., Holmes, A., 2000. Variability in fMRI: an examination of intersession differences. NeuroImage 11, 708–734, doi:10.1006/nimg.2000.0562.

McKeown, M., Makeig, S., Brown, G., Jung, T.-P., Kindermann, S., Sejnowski, T., 1998. Analysis of fMRI by blind separation into independent spatial components. Hum. Brain Mapp. 6 (3), 160–188.

Menon, R., Kim, S.-G., 1999. Spatial and temporal limits in cognitive neuroimaging with fMRI. Trends Cogn. Sci. 3 (6), 207–216.

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

Menon, R., Ogawa, S., Hu, X., Strupp, J., Andersen, P., Ugurbil, K., 1995. BOLD based functional MRI at 4 Tesla includes a capillary bed contribution: echo-planar imaging mirrors previous optical imaging using intrinsic signals. Magn. Reson. Med. 33, 453–459.

Miezin, F., Maccotta, L., Ollinger, J., Petersen, S., Buckner, R., 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, doi:10.1006/nimg.2000.0568.

Noll, D., Genovese, C., Nystrom, L., Forman, S., Eddy, W., Cohen, J., 1997. Estimating test-retest reliability in fMRI II: application to sensory-motor and cognitive activation. Magn. Reson. Med. 38, 508–517.

Norris, D., 2000. Reduced power multi-slice MDEFT imaging. J. Magn. Reson. Imag. 11, 445–451.

Ogawa, S., Lee, T.M., Kay, A.R., Tank, D.W., 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. USA 87, 9868–9872.

Rombouts, S., Barkhof, F., Hoogenraad, F., Sprenger, M., Scheltens, P., 1998. Within-subject reproducibility of visual activation patterns with functional magnetic resonance imaging using multislice echo planar imaging. Magn. Reson. Imag. 16 (2), 105–113.

Schacter, D., Buckner, R., Koutstaal, W., Dale, A., Rosen, B., 1997. Late onset of anterior prefrontal activity during true and false recognition: an event-related fMRI study. NeuroImage 6(4), 259–269.

Seber, G., Wild, C., 1989. Nonlinear Regression. Wiley, New York.

Stone, J., Porrill, J., Porter, N., Wilkinson, I.D., 2002. Spatiotemporal independent component analysis of event-related fMRI data using skewed probability density functions. NeuroImage 15, 407–421, doi:10.1006/nimg.2001.0986.

Stroop, J., 1995. Studies of inference in serial verbal reactions. J. Exp. Psychol. 18, 643–662.

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

Thierry, G., Boulanouar, K., Kherif, F., Ranjeva, J., Demonet, J., 1999. Temporal sorting of neural components underlying phonological processing. Neuroreport 10 (12), 2599–2603.

Ugurbil, K., Garwood, M., Hendrich, K., Hinke, R., Hu, X., Menon, R., Merkle, H., Ogawa, S., Salmi, R., 1993. Imaging at high magnetic fields: initial experiences at 4 Tesla. Magn. Reson. Quart. 9, 259–277.

Waldvogel, D., van Gelderen, P., Immisch, I., Pfeiffer, C., Hallett, M., 2000. The variability of serial fMRI data: correlation between a visual and a motor task. Neuroreport 11 (17), 3843–3847.

Worsley, K., Friston, K., 1995. Analysis of fMRI time-series revisited—again. NeuroImage 2, 173–181.

Zarahn, E., Aguirre, G., D'Esposito, M., 1997. Empirical analyses of BOLD fMRI statistics. NeuroImage 5, 179–197.

Zysset, S., Müller, K., Lohmann, G., von Cramon, D., 2001. Color-word matching stroop task: separating interference and response conflict. NeuroImage 13, 29–36, doi:10.1006/nimg.2000.0665.