ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

NeuroImage

NeuroImage 18 (2003) 798–805

www.elsevier.com/locate/ynimg

Technical Note

Estimating efficiency a priori: a comparison of blocked and randomized designs

Andrea Mechelli,[a,*] Cathy J. Price,[a] Rik N.A. Henson,[a,b] and Karl J. Friston[a]

[a] *Wellcome Department of Imaging Neuroscience, Institute of Neurology, 12 Queen Square, London WC1N 3BG, UK*
[b] *Institute of Cognitive Neuroscience, 17 Queen Square, London WC1N 3AR, UK*

Received 9 April 2002; revised 1 August 2002; accepted 7 August 2002

Abstract

This technical note deals with a priori estimation of efficiency of functional magnetic resonance imaging (fMRI) designs. The efficiency of an estimator is a measure of how reliable it is and depends on error variance (the variance not modeled by explanatory variables in the design matrix) and the design variance (a function of the explanatory variables and the contrast tested). Changes in the experimental design can induce changes in the variance of estimated responses. This translates into changes in the standard error of the response estimate or equivalently into changes in efficiency. One consequence is that statistics, testing for the same activation in different contexts (i.e., experimental designs), can change substantially even if the activation and error variance are exactly the same. We demonstrate this effect using an event-related fMRI study of single word reading during blocked and randomized trial presentations. Furthermore, we show that the error variance can change with the experimental design. This highlights a problem with a priori comparison of efficiency for two or more experimental designs, which usually assumes identical error variance.

Introduction

This article is about the a priori estimation of efficiency of functional magnetic resonance imaging (fMRI) designs. The efficiency of response estimation is a measure of the reliability with which model parameters are estimated and affects the sensitivity with which experimental effects are detected. In Friston et al. (1999), we presented a mathematical framework for computing the expected efficiency for various classes of experimental designs and compared contrasts testing for evoked responses per se and differential responses among trial types. We showed that the efficiency of a particular design depends on the contrast tested. Subsequently, there has been interest in the relative efficiency of contrasts testing for the amplitude of a response modeled by a single contrast or basis function (e.g., a canonical hemodynamic response function or HRF) and contrasts testing for a response modeled by multiple basis functions (e.g., a finite impulse response model). This distinction (Liu et al., 2001; Bandettini and Cox, 2000) has been referred to as "detection" versus "estimation" and reiterates the critical point that an efficient design for one contrast may not be optimal for another.

In this technical note, we revisit the relative efficiencies of different contrasts in the context of blocked and randomized designs in event-related experiments. Efficiency of response estimation is inversely related to the estimator variance (see Appendix). The estimator variance factorizes into the error variance and the design variance. The error variance (σ^2) is the residual variance after evoked changes, modeled by the design matrix, have been discounted. The design variance ($\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}$) is a function of the contrast ($\mathbf{c}$) and the design matrix ($\mathbf{X}$) and embodies the variance of the explanatory variables and the correlations among them. The design variance can be thought of as the estimator variance when $\sigma^2 = 1$.

In Friston et al. (1999), we showed that, in the context of response detection using a single basis function, blocked designs are more efficient than randomized designs. A useful way to understand why blocked presentation of trials is more efficient is in terms of the frequencies in which experimentally induced variance lies. Blocked designs typi-

* Corresponding author.

1053-8119/03/$ – see front matter
doi:10.1016/S1053-8119(02)00040-X

cally induce greater variance in low-frequency components that are "passed" by the HRF. This results in the standard error being smaller for contrasts testing for activations in blocked than randomized designs (Josephs and Henson, 1999; Paradis et al., 1998). The *t* values will therefore be higher in the blocked designs even if the amplitude of the hemodynamic responses and the parameter estimates ($\mathbf{c}^T \hat{\boldsymbol{\beta}}$) are identical. This leads to the somewhat counterintuitive prediction that significant effects can be evident in blocked but not randomized designs and yet the amplitude of the hemodynamic responses may not be significantly different. In a design that includes both blocked and randomized components, this means that it should be possible to show significant effects within the blocked but not the randomized component and yet no interaction between the effects and the experimental design. This is precisely what we show in the study below by quantifying the effects empirically and comparing the standard errors of blocked and randomized designs. A critical consequence of these observations is that one should never *compare statistics* to make inferences about differential responses. One should always make a *statistical comparison*, generally testing for an interaction. For instance, to assess the effect of stimulus sequence on hemodynamic responses, one should avoid anecdotal comparison of sequence-specific *Z* scores that could be confounded by differential efficiency. Rather, one should examine the interaction between hemodynamic response and stimulus sequence, by comparing directly hemodynamic responses for different stimulus sequences.

Our study also allowed us to address a second issue which relates to the use of the design variance and the error variance in a priori estimation of efficiency. The design variance (i) depends on the contrast and design matrix only, (ii) can be computed a priori, and (iii) is the same across the whole brain. In contrast, the error variance (i) depends on cognitive/ physiological effects (e.g, the hemodynamic responses may be more variable in one context relative to another), (ii) can only be estimated by performing a statistical analysis, and (iii) is voxel-specific. In other words, the relative efficiency of response estimation can only be quantified a priori by assuming that the error variance is independent of changes in the experimental design. For this assumption to be met, the responses need to conform to a linear convolution model, which embodies two further assumptions: nonlinearities can be discounted and the form of the hemodynamic response function is the same for different experimental designs. When these assumptions are violated, differences in error variance may arise, thereby compounding the relative efficiency of the designs. This study investigates whether the relative efficiency of blocked and randomized designs can be predicted under the assumptions adopted generally in the context of a priori estimation of efficiency. Differences in the error variance would indicate that both the error variance and the design need to be considered when estimating efficiency. Alternatively, if differences in error variance are negligible, the relative efficiency can be predicted from the design alone.

In summary, we present an event-related fMRI study of single word reading which involved acquiring data using two stimulus sequences associated with different efficiencies. We aimed to (i) show that differences in efficiency, attributable solely to experimental design, may lead to identical responses that can be detected in one presentation mode but not the other, and (ii) test whether differences in error variance were significant. If this is the case, both the error variance and the design need to be considered when estimating efficiency.

Methods

The study was approved by the National Hospital for Neurology and Institute of Neurology Medical Ethic's Committee.

Subjects

Informed consent was obtained from 12 right-handed volunteers (9 men), aged between 20 and 38 years (with a mean age of 26 years), with English as their first language.

Design

The stimuli were written words presented in lowercase (courier font) one at a time on a visual monitor with a *minimal* stimulus onset asynchrony (SOA) of 1.5 s. Subjects were instructed to fixate on a cross in the middle of the screen and read words silently as soon as they appeared. An eye tracker was used to monitor the eye movements of the subjects, to ensure that they kept their eyes open and scanned the stimuli. To establish an interstimulus baseline, null events were included with an occurrence probability of 0.25 (Josephs and Henson, 1999), producing a *mean* SOA of 2 s.

Two variables were manipulated, conforming to a 3×2 factorial design. The first was stimulus duration: words remained on the screen for 200, 600, or 1000 ms. This factor was manipulated within a scanning session. The second variable was presentation sequence: blocked versus randomized. In blocked sequences, stimuli were presented in trains of 35 of the same duration (block length 70 s). In randomized sequences, stimuli of different duration were randomly intermixed (see Fig. 1). The sequence factor was manipulated between scanning sessions.

Note that both sequences were "randomized" in the sense that the occurrence of stimuli conformed to a stationary stochastic design (Friston et al., 1999). The distinction between "blocked" and "randomized" was from the point of view of the ordering of stimulus types. In the blocked sequence, stimuli *of the same duration* were presented in blocks, whereas in the randomized sequence stimuli *of different durations* were intermixed.

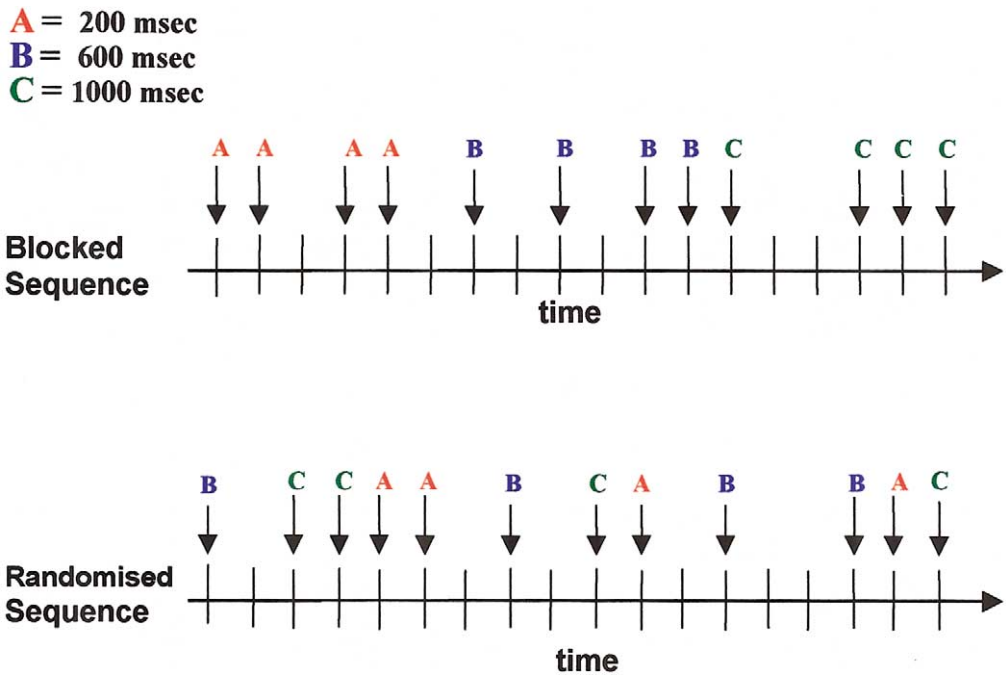

Fig. 1. Examples of blocked and randomized (i.e., event-related) presentations. In both blocked and randomized sequences, words remained on the screen for 200, 600, or 1000 ms. However, in the blocked sequence stimuli were presented in trains of 35 stimuli with the same duration, whereas in the randomized sequence stimuli of different durations were presented randomly. In both cases, null events occurred randomly to enable comparison of stimulus effects with interstimulus baseline (i.e., fixation).

Stimuli

Each subject was presented with 420 words composed of four, five, or six letters. These words, with regular grapheme-phoneme relationships, were allocated to 12 different sets with 35 words in each set, matched across set for frequency (Kucera and Francis, 1967), length, and number of syllables. Over subjects, each set of words was presented an equal number of times at each duration in both the blocked and the randomized sessions.

Scanning technique

A 2-T Siemens Vision system (Siemens, Erlangen, Germany) was used to acquire both T_1 anatomic volume images ($1 \times 1 \times 1.5$ mm voxels) and T_2*-weighted echoplanar images (64×64, 3×3 mm pixels, TE = 40 ms) with BOLD contrast. Each echoplanar image comprised 35 axial slices 1.8 mm thick with a 1.2-mm slice interval, giving a resolution of 3 mm. The repetition time (TR) was 3.15 s/volume. With a minimal SOA of 1.5 s, this results in sampling every 15 ms of peristimulus time (see Josephs et al., 1997; Price et al., 1999).

Data were acquired during four sessions, each comprising 74 volume images. The first six volumes in each session were discarded to allow for T_1 equilibration effects. In two of the four sessions stimulus duration was randomized. In the remaining two sessions, stimulus duration was blocked and counterbalanced across sessions (i.e., ABC in session 1 and CBA in session 2), that is, within and between subjects. Sessions with randomized and blocked sequences were also alternated in a counterbalanced order within and between subjects.

Preprocessing

Data were analyzed with statistical parametric mapping (SPM99; Wellcome Department of Imaging Neuroscience, London, UK; http//www.fil.ion.ucl.ac.uk/spm/), running under Matlab5.3 (Mathworks Inc., Sherbon, MA). All volumes from each subject were realigned using the first as reference and resliced using sine interpolation, adjusting for residual motion-related signal changes. To correct for different acquisition times, the signal measured in each slice was then shifted relative to the acquisition of the middle slice using sinc interpolation. Images were spatially normalized (Friston et al., 1995a) to a standard T_2* template in the space of Talairach and Tournoux (1988) using nonlinear-basis functions. This spatial transformation was also applied to the coregistered structural T_1 volume. Functional data were spatially smoothed with a 6-mm full-width at half-maximum isotropic Gaussian kernel, to compensate for residual variability after spatial normalization and to permit application of Gaussian random field theory for corrected statistical inference (Friston et al., 1995b).

Statistical analysis

The aim of the statistical analysis was to compare the effect of stimulus duration (1000 vs 200 ms) during the blocked and randomized sequences and to establish the contribution of noise and design to the detectability of these effects. Three statistical models were required. The first involved data from all four scanning sessions to compare the effect of stimulus duration during blocked and randomized sequences ("combined analysis"). The second and third statistical models involved independent analyses of blocked sequences or randomized sequences only. These "independent analyses" allowed us to estimate error variance for each type of stimulus sequence in regions showing effects in the combined analysis.

The data were high-pass filtered using a set of discrete cosine basis functions with a cutoff period of 512 s. Although this cutoff period only removed low-frequency drifts, it was chosen to preserve the experimental variance in both the blocked and the randomized design. The temporal autocorrelations in the errors were estimated using a restricted maximum likelihood (ReML) and a AR(1) + white noise model (Friston et al., 2002) and used to make the appropriate nonsphericity adjustment at the point of inference.

Combined analysis

Here we assumed that the error variance was the same for both blocked and randomized components of a compound design looking at the same activation. This was implicit in pooling the data in the same statistical model and assuming sphericity or homogeneity of error variance. This is an important issue from two perspectives: (i) the validity of the model and ensuing inference and (ii) the a priori estimation of efficiency.

Stimuli were classified into six event types according to presentation (blocked and randomized) and duration (200, 600, or 1000 ms). Both blocked and randomized sequences were modeled in the same "event-related" manner, in which responses to individual stimuli were modeled as brief bursts of neural/synaptic activity (delta functions) convolved by a HRF (Friston et al., 1998). The resulting functions were used as regressors in a general linear model. Although the stimuli ranged in duration, we did not model the width of the HRF since this was unlikely to differ measurably under the linear convolution model. The parameter estimates for the height of the canonical response for each event type were estimated using conventional least squares as implemented in the SPM99 software. Contrasts included (i) the effect of stimulus duration over both stimulus sequences, (ii) the effect of stimulus duration specific to either blocked or randomized presentation, and (iii) the interaction between duration and sequence. The statistical threshold for the effect of stimulus duration was set at 10 or more contiguous voxels surviving a height threshold of $P < 0.05$ (corrected for multiple comparisons). The threshold for the interaction between stimulus duration and presentation was relaxed ($P < 0.001$ uncorrected for multiple comparisons) to reduce type II errors. The maxima of the identified regions were localized on the normalized structural images and labeled using the nomenclature of Talairach and Tournoux (1988).

To fully characterize our data, the t statistic for each identified effect was decomposed into the contrast (e.g., the numerator of the t statistic) and standard error (e.g., the denominator of the t statistic) (see Appendix). The contrast depends on the value of the parameter estimates ($\hat{\boldsymbol{\beta}}$) and is an index of the size of the effect of interest. The standard error is inversely related to efficiency and can be factorized into the design and error variance. Since the error variance in the combined analysis is operationally the same for the blocked and randomized sequences, identifying differences in the error variance for the different sequences required separate statistical models.

Independent analyses

Within each mode of presentation (blocked or randomized), stimuli were classified into three event types according to the time they remained on the screen: 200, 600, or 1000 ms. Again the hemodynamic responses for each event-type were modeled with a canonical HRF (Friston et al., 1998) and the parameter estimates for the height of the canonical response for each event type were estimated using conventional least squares. Error variance from blocked and randomized data was compared using the F ratio (Sheskin, 1996), in maxima that showed positive effects of duration in the combined analysis. To maximize sensitivity to differences in the noise variance of the two data sets, comparisons were performed in this subset of voxels with a statistical threshold of 0.05 (uncorrected).

It should be noted that all analyses were performed in a fixed effect fashion, in which effects are averaged across subjects and compared to the within subject variability. In contrast, in a random effect analysis, the effect size is compared against the variability in the subject-specific parameter estimates, which embodies both the between subject and the within subject variability. This means that, given two studies that have the same between subject variability, the one with greater efficiency at the first level will also have greater efficiency at the second level. In this sense, fixed effect efficiency partially predicts random effect efficiency.

Results

Combined analysis

Main effects of stimulus duration

Positive effects of stimulus duration (1000 > 200 ms) across blocked and randomized presentation were found in

Table 1
Positive effect of stimulus duration

				Z scores			Contrast $\mathbf{c}^T \hat{\boldsymbol{\beta}}$		Standard error $(\sigma^2 \mathbf{c}^T (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c})^{1/2}$		Design variance $\mathbf{c}^T (\mathbf{X}^T\mathbf{X})^{-1}$		Error variance σ^2		Nonsphericity *P* value
				B and R	B	R	B	R	B	R	B	R	B	R	
Occipital															
R. superior occipital	16	−94	0	**7.5**	**7.7**	4.6	2.46	2.31	0.27	0.50	0.06	0.19	1.33	1.24	n.s.
L. superior occipital	−22	−96	16	**6.2**	**6.6**	3.5	1.78	1.75	0.27	0.50	0.06	0.19	1.25	1.35	n.s.
	−20	−100	8	**6.1**	**6.4**	3.4	1.58	1.56	0.25	0.45	0.06	0.19	1.04	1.10	n.s.
L. superior lingual	−12	−94	−8	**7.1**	**5.3**	**5.2**	1.47	2.68	0.28	0.52	0.06	0.19	1.47	1.25	<0.05 (B > R)
R. superior lingual	12	−94	−6	**6.6**	**5.6**	4.4	1.76	2.54	0.31	0.57	0.06	0.19	1.92	1.53	<0.05 (B > R)
L. inferior lingual	−4	−86	−14	4.3	**5.6**	3.1	2.33	2.40	0.42	0.76	0.06	0.19	2.64	3.05	<0.05 (R > B)
R. fusiform	30	−78	−16	**5.8**	**7.2**	2.3	2.65	1.45	0.34	0.62	0.06	0.19	1.97	2.18	<0.05 (R > B)
L. fusiform	−30	−78	−20	**5.6**	**7.4**	2.4	2.50	1.48	0.34	0.63	0.06	0.19	1.90	2.15	<0.05 (R > B)
Frontal															
R. superior frontal	18	66	24	3.1	**5.5**	0.5	1.92	0.31	0.34	0.63	0.06	0.19	1.28	1.44	<0.05 (R > B)

Note. Inferences were made at $P < 0.05$ (corrected for multiple comparisons) with an extent threshold of 10 or more contiguous voxels. *Z* scores significant at $P < 0.05$ (corrected) are reported in bold. While *Z* scores, standard errors and design variances for blocked (B) and randomized (R) data sets were derived from the combined analysis, the error variances were identified by performing independent analyses. Error variances for blocked and randomized data sets were compared with an *F* ratio at an uncorrected threshold of 0.05 (n.s., not significant). In this context, nonsphericity refers to the ratio of error variances for blocked and randomized designs. For simplicity, in the table headings, we do not include the autocorrelation matrix (i.e., Σ, see Appendix) in the standard error and design variance formulae. However, note that the variance estimators reported in this article used a form for the serial correlations that conforms to a AR(1) + white noise model (Friston et al., 2002).

the bilateral superior occipital, superior lingual, and fusiform gyrus (see Table 1 for details). These effects were associated with increased activation for reading at 1000 ms relative to rest ($P < 0.05$ corrected for multiple comparisons). There were no interactions between stimulus duration and sequence in any of these areas, even when the statistical threshold was lowered to $P < 0.01$ (uncorrected). Negative effects of stimulus duration (200 > 1000 ms) were not found.

It should be noted that there were other regions that showed a duration by sequence interaction. These universally showed a greater duration effect under blocked presentation. However, these results are not reported here because they indicate differences in the underlying response which precludes any comments on relative efficiency. Differences in the underlying response were hypothesized to arise from contextual effects (e.g., attentional set) in Price and Friston (1997) owing to different presentation scheduling. The methodologic implications of these set- or presentation-dependent responses will be discussed elsewhere. In this article, we focus on responses that were detected in the main effect of duration that, coincidentally, showed no context sensitivity in relation to presentation sequence.

Simple main effects of stimulus duration

A number of regions, including bilateral superior occipital, bilateral fusiform, right superior lingual, left inferior lingual, and right superior frontal gyrus, showed positive duration effects for the blocked mode ($P < 0.05$ corrected) with only trends ($P < 0.001$, uncorrected) for the randomized mode. For instance, in the left superior occipital gyrus, *Z* scores associated with the blocked sequence were 6.2 and 6.1, whereas *Z* scores associated with the randomized sequence were 3.5 and 3.4 (see Table 1). Bilateral fusiform and right superior frontal gyrus also showed positive duration effects for blocked presentation ($P < 0.05$ corrected) but not for randomized presentation even when lowering the threshold to $P < 0.001$ (uncorrected). For instance, in the right fusiform, the Z-score associated with the blocked mode was 7.2, whereas the *Z* score associated with randomized sequence was 2.3 (see Table 1). There were no duration by sequence interactions in any of these areas, even when the statistical threshold was lowered to $P < 0.01$ (uncorrected). There was no instance of a randomized effect being significant in the absence of a significant blocked effect.

Note that different realizations of our randomized design may have different efficiencies due to variability in both the positioning of null events and the ordering of stimuli (Dale, 1999). In contrast, different realizations of our blocked design will express less variability since they will vary only with respect to the positioning of null events. To address this point, we performed 1000 simulations to create a range of randomized and blocked designs. We found that the standard deviation of the design variances of randomized and blocked designs (0.047 and 0.004, respectively) was much smaller than the difference between randomized and blocked designs for the specific realizations used (i.e., 0.13; see Table 1). This indicates that the differential efficiency for the two presentation modes was not dependent on the specific realizations used in the present study.

Independent analyses

Independent analyses of blocked and randomized data sets were carried out solely to establish differences in the error variance for the different presentations. We found that error variance differed significantly between blocked and randomized data sets in a number of regions which showed experimental effects in the combined analysis. Specifically, error variance was larger for randomized than blocked presentation in bilateral fusiform, left inferior lingual, and right superior frontal gyrus. However, it was larger for blocked than randomized presentation in bilateral superior lingual areas (see Table 1 for details).

Discussion

As predicted theoretically, when the error variance is operationally fixed for blocked and randomized trial presentations, contrasts testing for blocked effects are more efficient leading to higher Z scores. The ratio of standard errors for the two contrasts is about 1:1.82. It should be noted that this ratio is the same for all regions. This is because, as discussed in the Introduction, the standard error depends on (i) the design variance (that is the same for each voxel) and (ii) the error variance (that is fixed for the blocked and the randomized model in the combined analysis). This means that the t (and Z equivalent) values are correspondingly higher for the simple effects of duration under blocked presentation. Although there is anecdotal evidence that the effects ($\hat{\boldsymbol{\beta}}$) reported in Table 1 may vary for randomized and blocked presentation in some regions, the absence of an interaction (even at $P < 0.01$ uncorrected) demonstrates they are not significantly different. One of the most telling results here is a Z score of 6.2 and 6.1 in the left superior occipital gyrus under blocked presentation and a Z score of 3.5 and 3.4 under randomized presentation. This disparity is a reflection of, and only of, the design efficiency enjoyed by blocked presentation (see Table 1). It has nothing to do with noise or, indeed, a different hemodynamic response. In short, although the responses were not significantly different in blocked and randomized presentation modes, activations could be detected at corrected level in blocked trials that could not be found in event-related trials.

Our results also indicate that the behavior of error variance may depend on the experimental design. This suggests that, while it is possible to use the design variance to estimate the impact of the experimental design on efficiency a priori, it is not possible to assume that the impact of error variance on efficiency is negligible. This means that both the design and the error variance must be taken into account when estimating the relative efficiency of blocked and randomized modes. Our findings may result from the violation of any of the assumptions underlying a priori comparison of efficiency. For instance, differences in the shape of the hemodynamic response or expression of BOLD nonlinearities may have led to differential error variance for the blocked and the randomized presentation.

The impact of the experimental design on error variance was variable across the brain. While most regions in the occipital cortex showed greater error variance for randomized than blocked presentation, the bilateral superior lingual gyrus showed greater error variance for blocked than randomized presentation. Furthermore, early visual areas including the bilateral occipital gyrus did not show differential error variance for the two presentation modes. Such variability may speak to functional specialization. For instance, in areas involved in high-order functions, cognitive effects may affect the shape and timing of the neuronal response, which may result in differential error variance estimates for blocked and randomized sequence. One possibility is that, in areas involved in high-order functions, there is a greater neuronal response for the stimulus immediately following a change in duration. Because such duration transitions are more common in the randomized than blocked design, this unmodeled effect would engender error. In contrast, in early visual areas such as the superior occipital gyrus, neuronal activity may simply reflect the amount of overall stimulation while not being affected by variables such as stimulus sequence.

Conclusion

We have shown that the relative efficiency of different presentation modes is directly expressed as differences in standard error that can have substantial effects on the ensuing statistics. When assessing responses in different experimental contexts, it is therefore important to examine the response by context interaction and avoid anecdotal comparison of the contrast-specific statistics. Furthermore, our results indicate that the error variance may depend on the experimental design. This makes a priori estimation of efficiency problematic: both the design and error variance must be taken into account but the latter can only be estimated by performing a statistical analysis.

Acknowledgments

This work was funded by the Wellcome Trust. We thank two anonymous reviewers for their helpful suggestions.

Appendix

The general linear model and efficiency

In the general linear model applied to fMRI time series, the response variable y is expressed in terms of a linear combination of explanatory variables in a design matrix $\mathbf{X}$

and a normally distributed error term $N(0, \sigma^2\Sigma)$. For simplicity we will assume that the errors are independent, that is, $\Sigma = \mathbf{I}$, although the variance estimators reported in this article used a form for the serial correlations that conforms to a AR(1) + white noise model (Friston et al., 2002). See Worsley and Friston (1995) for a treatment of the degrees of freedom that uses the Satterthwaite approximation to accommodate serial correlations when $\Sigma \neq \mathbf{I}$.

$$y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

$$\langle y \rangle = \mathbf{X}\boldsymbol{\beta} \tag{1a}$$

$$\mathrm{var}(y) = \sigma^2\Sigma. \tag{1b}$$

Ordinary least squares parameter estimates $\hat{\boldsymbol{\beta}}$ obtain using the pseudoinverse (denoted by $+$) of the design matrix

$$\hat{\boldsymbol{\beta}} = \mathbf{X}^{+}\mathbf{y} \tag{2}$$

$$\langle \hat{\boldsymbol{\beta}} \rangle = \mathbf{X}^{+}\langle \mathbf{y} \rangle = \boldsymbol{\beta} \tag{2a}$$

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) = \mathbf{X}^{+}\,\mathrm{var(y)}\mathbf{X}^{+\mathrm{T}} = \sigma^2\mathbf{X}^{+}\Sigma\mathbf{X}^{+\mathrm{T}} = \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}. \tag{2b}$$

Inferences about effects of interest can be made using t statistics. An effect of interest is specified by a vector of contrast weights $\mathbf{c}$ that gives a weighted sum or compound of parameter estimates $\mathbf{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}}$, referred to as a contrast. The contrasts of parameter estimates $\mathrm{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}}$ are an index of the size of the effects of interest. This is because the expected parameter estimates $\langle \hat{\boldsymbol{\beta}} \rangle$ are equal to the true parameters $\boldsymbol{\beta}$ (see Eq. (2a)). The t statistic is simply the contrast divided by its estimated standard error, e.g., the square root of the contrast variance:

$$t = \mathbf{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}}/(\mathrm{var}(\mathbf{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}}))^{1/2}, \tag{3}$$

where the contrast variance $\mathrm{var}(\mathbf{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}})$ is a function of error variance σ^2 and the design matrix $\mathbf{X}$. From Eq. (2b):

$$\mathrm{var}(\mathbf{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{c}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{c}. \tag{4}$$

In the present article, we refer to $\mathbf{c}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{c}$ as design variance in contradistinction to $\sigma^2\mathbf{c}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{c}$, which is contrast or estimator variance (note that when $\sigma^2 = 1$, the design and contrast variances are equivalent).

The contrast variance can be used to estimate the efficiency of an estimator for a specified contrast of interest c. Specifically, the efficiency of an estimator is inversely related to the contrast variance; for example, it decreases with error variance σ^2 and design variance $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$:

$$(\text{Standard error})^2 = \frac{1}{\text{Efficiency}} = \sigma^2\mathbf{c}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{c}. \tag{5}$$

Here efficiency can be regarded as a special case of the Cramer-Frechet formulation

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) \geq \left(1 + \frac{\partial \mathrm{b}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right) I(\boldsymbol{\beta})^{-1} \left(1 + \frac{\partial \mathrm{b}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)^{\mathrm{T}},$$

which sets a minimum bound on the error variance of any estimator of the parameters (where $b(\boldsymbol{\beta})$ is the estimator bias and $I(\boldsymbol{\beta})$ is the information matrix). Specifically, we assume (i) i.i.d. Gaussian noise—that is, $I(\boldsymbol{\beta}) = (\mathbf{X}^{\mathrm{T}}\mathbf{X}/\sigma^2)$; and (ii) no bias—$b(\boldsymbol{\beta}) = 0$.

When contrast matrices are specified (e.g., when a number of basis functions are equally interesting), efficiency can be computed by using the trace operator (Dale, 1999):

$$\frac{1}{\text{Efficiency}} \propto \mathrm{trace}\,[\mathbf{c}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{c}]. \tag{6}$$

Clearly to compute the t statistic, standard error or efficiency one must estimate σ^2. This is achieved through dividing the residual sum of squares by the degrees of freedom using

$$\langle \mathbf{r}^{\mathrm{T}}\mathbf{r} \rangle = \langle \mathrm{trace}(\mathbf{r}\mathbf{r}^{\mathrm{T}}) \rangle = \langle \mathrm{trace}(\mathbf{R}\mathbf{y}\mathbf{y}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}) \rangle = \sigma^2\,\mathrm{trace}(\mathbf{R}\mathbf{R}^{\mathrm{T}}) \tag{7}$$

$$\sigma^2 \approx \frac{\mathbf{r}^{\mathrm{T}}\mathbf{r}}{\mathrm{trace}(\mathbf{R}\mathbf{R}^{\mathrm{T}})} = \frac{\mathbf{r}^{\mathrm{T}}\mathbf{r}}{\mathrm{trace}(\mathbf{R})}, \tag{7a}$$

where $\mathbf{r}$ is the vector of residuals and $\mathbf{R}$ is the residualforming matrix:

$$\mathbf{r} = \mathbf{R}\mathbf{y} \tag{7b}$$

$$\mathbf{R} = \mathbf{I} - \mathbf{X}\mathbf{X}^{\mathrm{T}}. \tag{7c}$$

References

Bandettini, P.A., Cox, R.W., 2000. Event-related fMRI contrast when using constant interstimulus interval: theory and experiment. Magn. Reson. Med. 43, 540–548.

Dale, A.M., 1999. Optimal experimental design for event-related fMRI. Hum. Brain Map. 8, 109–114.

Friston, K.J., Ashburner, J., Frith, C.D., Poline, J.-B., Heather, J.D., Frackowiak, R.S.J., 1995a. Spatial registration and normalization of images. Hum. Brain Map. 2, 1–25.

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

Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., Turner, R., 1988. Event-related fMRI: characterizing differential responses. NeuroImage 7, 30–40.

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

Friston, K.J., Glaser, D.E., Henson, R.N.A., Kiebel, S., Phillips, C., Ashburner, J., 2002. Classical and Bayesian inference in neuroimaging: applications. NeuroImage 16, 484–512.

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

Josephs, O., Henson, R.N.A., 1999. Event-related fMRI: modelling, inference and optimisation. Phil. Trans. R. Soc. Lond. 354, 1215–1228.

Kucera, H., Francis, W.H. 1967. Computational Analysis of Present-Day American English. Brown Univ. Press, Providence, RI.

Liu, T.T., Frank, L.R., Wong, E.C., Buxton, R.B., 2001. Detection power, estimation efficiency, and predictability in event-related fMRI. NeuroImage 13, 759–773.

Paradis, A.-L., Van de Morrtele, P.-F., Le Bihan, D., Poline, J.-B., 1998. Do high temporal frequencies of the event-related fMRI response have a more specific spatial localization. NeuroImage 7, S606.

Price, C.J., Friston, K.J., 1997. The temporal dynamics of reading: a PET study. Proc. R. Soc. Lond. B 264, 1785–1791.

Price, C.J., Veltman, D., Ashburner, J., Josephs, O., Friston, K., 1999. The critical relationship between the timing of stimulus presentation and data acquisition in fMRI. NeuroImage 10, 36–45.

Sheskin, D.J., 1996. The Handbook of Parametric and Nonparametric Statistical Procedures. CRC Press, Boca Raton, FL.

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

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