NeuroImage 125 (2016) 756–766

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Effect of trial-to-trial variability on optimal event-related fMRI design: Implications for Beta-series correlation and multi-voxel pattern analysis

Hunar Abdulrahman [a,b], Richard N. Henson [a,*]

[a] MRC Cognition & Brain Sciences Unit, Cambridge, England, United Kingdom
[b] University of Cambridge, Cambridge, United Kingdom

ARTICLE INFO

Article history:
Received 20 July 2015
Accepted 3 November 2015
Available online 6 November 2015

Keywords:
fMRI design
General Linear Model
Bold variability
Least squares all
Least squares separate
MVPA
Trial based correlations

ABSTRACT

Functional magnetic resonance imaging (fMRI) studies typically employ rapid, event-related designs for behavioral reasons and for reasons associated with statistical efficiency. Efficiency is calculated from the precision of the parameters (Betas) estimated from a General Linear Model (GLM) in which trial onsets are convolved with a Hemodynamic Response Function (HRF). However, previous calculations of efficiency have ignored likely variability in the neural response from trial to trial, for example due to attentional fluctuations, or different stimuli across trials. Here we compare three GLMs in their efficiency for estimating average and individual Betas across trials as a function of trial variability, scan noise and Stimulus Onset Asynchrony (SOA): "Least Squares All" (LSA), "Least Squares Separate" (LSS) and "Least Squares Unitary" (LSU). Estimation of responses to individual trials in particular is important for both functional connectivity using "Beta-series correlation" and "multi-voxel pattern analysis" (MVPA). Our simulations show that the ratio of trial-to-trial variability to scan noise impacts both the optimal SOA and optimal GLM, especially for short SOAs < 5 s: LSA is better when this ratio is high, whereas LSS and LSU are better when the ratio is low. For MVPA, the consistency across voxels of trial variability and of scan noise is also critical. These findings not only have important implications for design of experiments using Beta-series regression and MVPA, but also statistical parametric mapping studies that seek only efficient estimation of the mean response across trials.

Introduction

Many fMRI experiments use rapid presentation of trials of different types (conditions). Because the time between trial onsets (or Stimulus Onset Asynchrony, SOA) is typically less than the duration of the BOLD impulse response, the responses to successive trials overlap. The majority of fMRI analyses use linear convolution models like the General Linear Model (GLM) to extract estimates of responses to different trial-types (i.e., to deconvolve the fMRI response; Friston et al., 1998). The parameters of the GLM, reflecting the mean response to each trial-type, or even to each individual trial, are estimated by minimizing the squared error across scans (where scans are typically acquired with repetition time, or TR, of 1–2 s) between the timeseries recorded in each voxel and the timeseries that is predicted, based on i) the known trial onsets, ii) assumptions about the shape of the BOLD impulse response and iii) assumptions about noise in the fMRI data.

Many papers have considered how to optimize the design of fMRI experiments, in order to maximize statistical efficiency for a particular contrast of trial-types (e.g., Dale, 1999; Friston et al., 1999; Josephs and Henson, 1999). However, these papers have tended to consider only the choice of SOA, the probability of occurrence of trials of each type and the modeling of the BOLD response in terms of a Hemodynamic Response Function (HRF) (Henson, 2015; Liu et al., 2001). Few studies have considered the effects of variability in the amplitude of neural activity evoked from trial to trial (though see Josephs and Henson, 1999; Duann et al., 2002; Mumford et al., 2012). Such variability across trials might include systematic differences between the stimuli presented on each trial (Davis et al., 2014). This is the type of variability, when expressed differently across voxels, that is relevant to multi-voxel pattern analysis (MVPA), such as representational similarity analysis (RSA) (Mur et al., 2009). However, trial-to-trial variability is also likely to include other components such as random fluctuations in attention to stimuli, or variations in endogenous (e.g., pre-stimulus) brain activity that modulates stimulus-evoked responses (Becker et al., 2011; Birn, 2007; Fox et al., 2006); variability that can occur even for replications of exactly the same stimulus across trials. This is the type of variability utilized by trial-based measures of functional connectivity between voxels (so-called "Beta-series" regression, Rissman et al., 2004).

If one allows for variability in the response across trials of the same type, then one has several options for how to estimate those responses

* Corresponding author at: MRC Cognition & Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom.
E-mail address: rik.henson@mrc-cbu.cam.ac.uk (R.N. Henson).

http://dx.doi.org/10.1016/j.neuroimage.2015.11.009

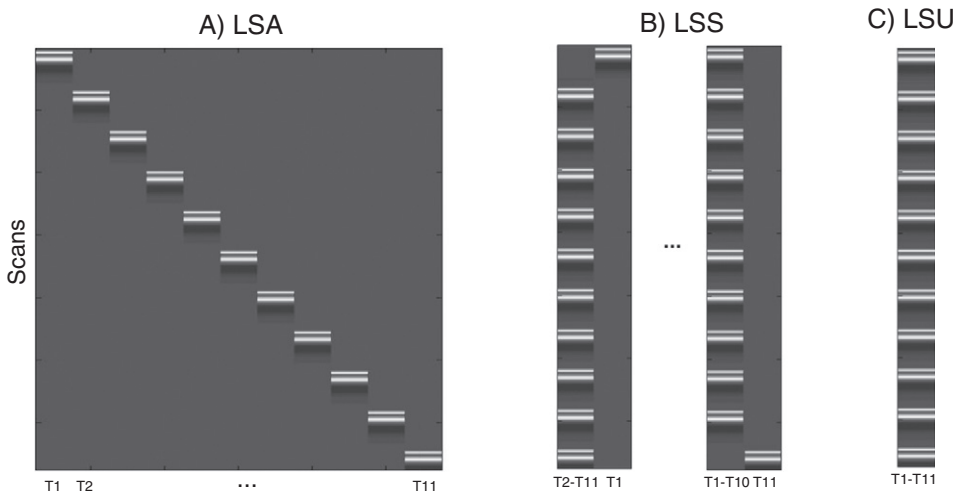

Fig. 1. Design matrices for (A) LSA (Least Squares-All), (B) LSS (Least Squares-Separate) and (C) LSU (Least Squares-Unitary). T(number) = Trial number.

within the GLM. Provided one has more scans than trials (i.e. the SOA is longer than the TR), and provided the HRF is modeled with single (canonical) shape (i.e., with one degree of freedom), one could model each trial as a separate regressor in the GLM (Fig. 1A). Mumford et al. (2012) called this approach "Least-Squares All" (LSA), in terms of the GLM minimizing the squared error across all regressors. Turner (2010) introduced an alternative called "Least-Squares Separate" (LSS; Fig. 1B). This method actually estimates a separate GLM for each trial. Within each GLM, the trial of interest (target trial) is modeled as one regressor, and all the other (non-target) trials are collapsed into another regressor. This approach has been promoted for designs with short SOAs, when there is a high level of collinearity between BOLD responses to successive trials (Mumford et al., 2012). For completeness, we also consider the more typical GLM in which all trials of the same type are collapsed into the same regressor, and call this model "Least-Squares Unitary" (LSU). Though LSU models do not distinguish different trials of the same type (and so trial variability is relegated to the GLM error term), they are used to estimate the mean response for each trial-type, and we show below that the precision of this estimate is also affected by the ratio of trial variability to scan noise.

In the current study, we simulated the effects of different levels of trial-to-trial variability, as well as scan-to-scan noise (i.e., noise), on the ability to estimate responses to individual trials, across a range of SOAs (assuming that neural activity evoked by each trial was brief – i.e., less than 1 s – and locked to the trial onset, so that it can be effectively modeled as a delta function). More specifically, we compared the relative efficiency of the three types of GLM – LSU, LSA and LSS – for three distinct questions: 1) estimating the population or sample mean of responses across trials, as relevant, for example, to univariate analysis of a single voxel (e.g., statistical parametric mapping), 2) estimating the response to each individual trial, as relevant, for example, to trial-based measures of functional connectivity between voxels (Rissman et al., 2004), and 3) estimating the pattern of responses across voxels for each trial, as relevant to MVPA (e.g., Mumford et al., 2012). In short, we show that different GLMs are optimal for different questions, depending on the SOA and the ratio of trial variability to scan noise.

Methods

We simulated fMRI timeseries for a fixed scanning duration of 45 min (typical of fMRI experiments), sampled every TR = 1 s. We modeled events by delta functions that were spaced with SOAs in steps of 1 s from 2 s to 24 s, and convolved with SPM's (www.fil.ion.ucl.ac.uk/spm) canonical HRF, scaled to have peak height of 1. The scaling of the delta-functions (true parameters) for the first trial-type (at a single voxel) was drawn from a Gaussian distribution with a population mean of 3 and standard deviation (SD) that was one of 0, 0.5, 0.8, 1.6, or 3. Independent zero-mean Gaussian noise was then added to each TR, with SD of 0.5, 0.8, 1.6 or 3,[1] i.e., producing amplitude SNRs of 6, 3.8, 1.9 or 1 respectively. (Note that, as our simulations below show, the absolute values of these standard deviations matter little; what matters is the ratio of trial variability relative to scan noise.)

For the simulations with two trial-types, the second trial-type had a population mean of 5. The two trial-types were randomly intermixed. For the simulations of two trial-types across two voxels, either the same sample of parameter values was used for each voxel (coherent trial variability), or different samples were drawn independently for each voxel (incoherent trial variability). The GLM parameters ("Betas", $\boldsymbol{\beta}$) were estimated by least-squares fit of each of the GLMs in Fig. 1:

$$\hat{\boldsymbol{\beta}}_{OLS} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

where $\mathbf{X}^T$ is the transpose of the GLM design matrix and $\mathbf{y}$ is a vector of fMRI data for a single voxel. In extra simulations, we also examined a L2-regularized estimator for LSA models (equivalent to ridge regression; see also Mumford et al., 2012):

$$\hat{\boldsymbol{\beta}}_{RLS} = \left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

where $\mathbf{I}$ is a scan-by-scan identity matrix and λ is the degree of regularization, as described in the Discussion section. A final constant term was added to remove the mean BOLD response (given that the absolute value of the BOLD signal is arbitrary). The precision of these parameter estimates was estimated by repeating the data generation and model fitting N = 10,000 times. This precision can be defined in several ways, depending on the question, as detailed in the Results section. Note that for regularized estimators, there is also a bias (whose trade-off with efficiency depends on the degree of regularization), tending to shrink the parameter estimates towards zero, but we do not consider this bias here.

Note that we are only considering the accuracy of the parameter estimates across multiple realizations (simulations, e.g., sessions, participants, or experiments), e.g., for a "random-effects" group analysis across participants. We do not consider the statistical significance (e.g., T-values) for a single realization, e.g., for a "fixed effects" within-participant analysis. The latter will also depend on the nature of the

[1] Note that in the special case of zero trial variability and zero scan noise, all parameters would be estimated perfectly, and so all GLMs are equivalent.