NeuroImage 147 (2017) 390–408

Contents lists available at ScienceDirect

NeuroImage

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

Comparison of fMRI analysis methods for heterogeneous BOLD responses in block design studies

Jia Liu[a,1], Ben A. Duffy[a,1], David Bernal-Casas[a], Zhongnan Fang[a,b], Jin Hyung Lee[a,b,c,d,*]

[a] Department of Neurology & Neurological Sciences, Stanford University, 1201 Welch Road, #P206, Stanford, CA, 94305 USA
[b] Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA
[c] Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
[d] Department of Neurosurgery, Stanford University, Stanford, CA 94305, USA

ARTICLE INFO

Keywords:
Hemodynamic response
Block design
fMRI data analysis
Optogenetic fMRI

ABSTRACT

A large number of fMRI studies have shown that the temporal dynamics of evoked BOLD responses can be highly heterogeneous. Failing to model heterogeneous responses in statistical analysis can lead to significant errors in signal detection and characterization and alter the neurobiological interpretation. However, to date it is not clear that, out of a large number of options, which methods are robust against variability in the temporal dynamics of BOLD responses in block-design studies. Here, we used rodent optogenetic fMRI data with heterogeneous BOLD responses and simulations guided by experimental data as a means to investigate different analysis methods' performance against heterogeneous BOLD responses. Evaluations are carried out within the general linear model (GLM) framework and consist of standard basis sets as well as independent component analysis (ICA). Analyses show that, in the presence of heterogeneous BOLD responses, conventionally used GLM with a canonical basis set leads to considerable errors in the detection and characterization of BOLD responses. Our results suggest that the 3rd and 4th order gamma basis sets, the 7th to 9th order finite impulse response (FIR) basis sets, the 5th to 9th order B-spline basis sets, and the 2nd to 5th order Fourier basis sets are optimal for good balance between detection and characterization, while the 1st order Fourier basis set (coherence analysis) used in our earlier studies show good detection capability. ICA has mostly good detection and characterization capabilities, but detects a large volume of spurious activation with the control fMRI data.

1. Introduction

Reliable detection of evoked blood oxygenation level dependent (BOLD) responses is critical to estimate the brain activation maps in fMRI studies. In addition, there has been an increasing interest in characterizing temporal features such as onset and duration to investigate activation timing of BOLD responses across brain regions and experimental conditions (Byers et al., 2015; Handwerker et al., 2012; Lindquist et al., 2009; Liu et al., 2015; Weitz et al., 2014). However, accurate detection and characterization remain challenging in scenarios where BOLD responses exhibit a large variability in the temporal dynamics (Aguirre et al., 1998; Gonzalez-Castillo et al., 2012; Handwerker et al., 2004), such as in studies of disease states (Amemiya et al., 2012; Matthews et al., 2006), and in small animal studies with anesthesia (Schlegel et al., 2015; Schroeter et al., 2014; Williams et al., 2010). In these cases, commonly used general linear model (GLM) (Friston et al., 1994) with a canonical hemodynamic response function (HRF) is often not the best choice. For example, in an fMRI study of motor control in human ischemic patients, GLM with a canonical HRF failed to detect motor cortex activation (Amemiya et al., 2012). It also failed to estimate temporal features of the BOLD responses (Calhoun et al., 2004; Lindquist et al., 2009). In these studies, onset and duration differences between experimental conditions were misinterpreted as differences in the amplitudes of evoked BOLD responses. These substantial detection and characterization errors stress the importance of proper choice of analysis methods.

Nevertheless, it is currently not clear which methods are optimal in scenarios of heterogeneous BOLD responses. This is partially due to the large set of analysis approaches available, yet few comprehensive evaluations have been conducted, especially in block-design studies. Over the past decades, dozens of methods have been proposed. Among the most accessible ones are those implemented in widely available

*Abbreviations:*ofMRI, optogenetic functional magnetic resonance imaging

* Corresponding author at: 1201 Welch Road, #P206, Stanford, CA 94305, USA.
E-mail address: ljinhy@stanford.edu (J.H. Lee).

[1] These authors contributed equally.

http://dx.doi.org/10.1016/j.neuroimage.2016.12.045
Received 20 September 2016; Accepted 15 December 2016
Available online 16 December 2016
1053-8119/

software packages, such as GLM with the canonical basis set (Calhoun et al., 2004; Friston et al., 1998; Henson et al., 2002; Steffener et al., 2010), the gamma basis set, the Fourier basis set, the finite impulse response (FIR) basis set, and the B-spline basis set (Genovese, 2000). Likewise, optimized methods for specific datasets have been considered. For example, colleagues have developed specific basis sets to estimate onset delays (Liao et al., 2002), implemented transient plus sustained models to detect transient responses in block-design experiments (Giraud et al., 2000; Harms and Melcher, 2002; Seifritz et al., 2002), and designed basis sets that incorporate prior information of BOLD responses (Woolrich et al., 2004). Additionally, data-driven methods are employed as they place few assumptions on the hemodynamic responses. Commonly used methods include independent component analysis (ICA) (Beckmann and Smith, 2004; Esposito et al., 2002; McKeown et al., 1998a; McKeown et al., 1998b), principal component analysis (PCA) (Backfrieder et al., 1996; Sychra et al., 1994), and fuzzy clustering analysis (Baumgartner et al., 2000; Chuang et al., 1999; Wismüller et al., 2002).

In block-design studies, only data-driven methods, such as ICA, PCA, and unsupervised clustering, have been compared on their detection and characterization performance (Baumgartner et al., 2000; Erhardt et al., 2011; Meyer-Baese et al., 2004), but not the more widely-used model-based approaches. In contrast, another study assessed several HRF models' ability to estimate HRF parameters from a block-design experiment, but did not examine detection performance (Shan et al., 2014). More often, comparisons were not conducted as the main purpose of the study, but to support the introduction of new approaches to analyze fMRI data (Calhoun et al., 2001a; Harms and Melcher, 2003; McKeown et al., 1998b; Moritz et al., 2003), or to highlight the heterogeneity of the observed BOLD responses (Amemiya et al., 2012; Gonzalez-Castillo et al., 2012; Pujol et al., 2009; Schlegel et al., 2015). As a result, it is difficult to derive a comprehensive evaluation from the literature, due to the limited range of statistical methods employed and/or assessment conducted in each study.

Here, we investigate the robustness of six widely available methods against heterogeneous BOLD responses in block-design studies. Given the fact that the vast majority of methods already incorporate information about the shape of evoked hemodynamic responses during the detection stage, we focused not only on each method's detection performance, but also on their characterization power (Degras and Lindquist, 2014; Makni et al., 2008). A detailed comparison of state-of-the-art methods for analyses of heterogeneous BOLD responses is presented. Evaluations are carried out in the GLM framework and include standard basis sets as well as ICA. In order to evaluate each methods' performance against fMRI data with heterogeneous BOLD responses, we use data from a recently published optogenetic fMRI (ofMRI) study of dynamic control of forebrain by central thalamus (Liu et al., 2015). To further validate each method's performance, we also use simulated data with varying temporal dynamics. Advantages and shortcomings of each approach are quantified using receiver operating characteristic (ROC) analysis and root-mean-square error (RMSE) of fit. Together, our results aim to provide practical recommendations on proper methods selection for analyzing block-design fMRI data with heterogeneous BOLD responses.

2. Methods

2.1. fMRI analysis methods

In this study, a set of six different approaches including model-based and data-driven methods was evaluated. The same block-design paradigm was used across methods. It consisted of 30 s baseline, followed by six 60 s cycles, each consisting of 20 s stimulation and 40 s rest, unless otherwise noted. To enable comparison across methods, a single statistical analysis platform is needed. Therefore, the linear regression platform in Statistical Parametric Mapping (SPM, Wellcome Trust Center for Neuroimaging) was employed for statistical analysis. All methods were evaluated by using different sets of regressors within the same GLM framework. The detailed description of each method is included as follows:

(i) The canonical basis set was selected from the SPM toolbox as one of the most commonly used methods. Model orders up to 3 were included in the evaluation. In the present study, GLM with a single canonical HRF as basis function is referred to as the 1st order canonical basis set. GLM with a canonical HRF and its temporal derivative as basis functions is referred to as the 2nd order canonical basis set. GLM with a canonical HRF and its temporal and dispersion derivatives as basis functions is referred to as the 3rd order canonical basis set. The canonical basis functions were first convolved with the stimulation paradigm before being used as regressors for the canonical basis set.

(ii) The gamma basis set was selected from the SPM toolbox as another widely available method. Model orders up to 4 were investigated. Each order includes a set of K gamma functions of increasing dispersions as basis functions, where K denotes the model order. Similar with the canonical basis set, the gamma basis functions were first convolved with the stimulation paradigm before being used as regressors for the gamma basis set.

(iii) The FIR basis set was included as one of the most flexible basis sets. The model order of 3 to 10 was investigated. Each order includes a set of K contiguous boxcar functions, in which the bin width of each boxcar function equals T/K, where K denotes the model order, and T represents the length of each stimulation cycle (60 s). For simplicity, only results from the odd numbers (e.g., model order of 3, 5, 7, and 9) are shown in figures. Additionally, we investigated the model order of 20, in which the bin width of each boxcar function equals our image acquisition interval (3 s), a common practice when employing the FIR basis set. Unlike the canonical and gamma basis sets, the FIR basis set was not convolved with the stimulation paradigm before being used as regressors.

(iv) The B-spline basis set was selected as another popular analysis method (Genovese, 2000; Schlegel et al., 2015). The model order of 3 to 10 was included in the evaluation. Each order includes a set of K cubic spline functions created using the program 3dDeconvolve in the AFNI software package (Cox, 1996; Ward, 2006), where K denotes the model order. Similar with the FIR basis set, only results from the odd numbers are shown for simplicity (e.g., model order of 3, 5, 7, and 9), and the B-spline basis set was not convolved with the stimulation paradigm before being used as regressors.

(v) The Fourier basis set was selected due to its capability to exploit the periodic nature of the experimental paradigm and evoked responses (Bullmore et al., 1996; Pinto et al., 2016). Model orders up to 5 were investigated. Each order includes a set of K sine and K cosine functions at harmonic frequencies: f_1, $2f_1$, ..., Kf_1 Hz, where K denotes the model order, and f_1 represents the frequency of repeated stimulation cycles (1/60 Hz). Similar with the FIR and B-spline basis sets, the Fourier basis set was not convolved with the stimulation paradigm before being used as regressors.

It is worth noting that, GLM with the 1st order Fourier basis set is mathematically equivalent with coherence analysis, a frequency-domain analysis method that is widely used in periodic block-design studies (Amemiya et al., 2012; Bandettini et al., 1993; Engel et al., 1997; Lee et al., 2010), including the ofMRI datasets we utilized in the present study (Liu et al., 2015). A coherence value was defined as a ratio of the magnitude of each time series' Fourier transform (F) at the frequency of repeated stimulation cycles (f_1, 1/60 Hz) and the total