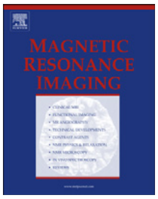
Magnetic Resonance Imaging 32 (2014) 721–735

Contents lists available at ScienceDirect

Magnetic Resonance Imaging

journal homepage: www.mrijournal.com



Deconvolution filtering: Temporal smoothing revisited

Keith Bush [a,]*, Josh Cisler [b]

[a] Department of Computer Science, University of Arkansas at Little Rock (UALR), Little Rock, AR 72204, USA
[b] Brain Imaging Research Center, University of Arkansas for Medical Sciences (UAMS), Little Rock, AR 72205, USA

ARTICLE INFO

Article history:
Received 12 August 2013
Revised 20 February 2014
Accepted 7 March 2014

Keywords:
Deconvolution
Filtering
fMRI
Imaging analysis
BOLD signal
Functional connectivity

ABSTRACT

Inferences made from analysis of BOLD data regarding neural processes are potentially confounded by multiple competing sources: cardiac and respiratory signals, thermal effects, scanner drift, and motion-induced signal intensity changes. To address this problem, we propose deconvolution filtering, a process of systematically deconvolving and reconvolving the BOLD signal via the hemodynamic response function such that the resultant signal is composed of maximally likely neural and neurovascular signals. To test the validity of this approach, we compared the accuracy of BOLD signal variants (i.e., unfiltered, deconvolution filtered, band-pass filtered, and optimized band-pass filtered BOLD signals) in identifying useful properties of highly confounded, simulated BOLD data: (1) reconstructing the true, unconfounded BOLD signal, (2) correlation with the true, unconfounded BOLD signal, and (3) reconstructing the true functional connectivity of a three-node neural system. We also tested this approach by detecting task activation in BOLD data recorded from healthy adolescent girls (control) during an emotion processing task.

Results for the estimation of functional connectivity of simulated BOLD data demonstrated that analysis (via standard estimation methods) using deconvolution filtered BOLD data achieved superior performance to analysis performed using unfiltered BOLD data and was statistically similar to well-tuned band-pass filtered BOLD data. Contrary to band-pass filtering, however, deconvolution filtering is built upon physiological arguments and has the potential, at low TR, to match the performance of an optimal band-pass filter. The results from task estimation on real BOLD data suggest that deconvolution filtering provides superior or equivalent detection of task activations relative to comparable analyses on unfiltered signals and also provides decreased variance over the estimate. In turn, these results suggest that standard preprocessing of the BOLD signal ignores significant sources of noise that can be effectively removed without damaging the underlying signal.

1. Introduction

The need to account for confounding factors in fMRI is a well-known problem in fMRI analysis that uses BOLD signals. The essential problem is that a BOLD signal is multidetermined: it is a mixture of the neurovascular consequences of neural firing combined with confounding factors, such as head motion, scanner artifact and signal drift, neurophysiological variability, etc. Consequently, signal fluctuations cannot be solely attributed to neural causes, which limits inferences about neural processes in functional imaging studies. Accordingly, attempts must be made to account for and/or remove confounding (i.e., non-neural in origin) sources of variance from BOLD data in order to foster more precise inferences to inform cognitive and clinical neuroscience.

Attempts to address this problem date back to seminal papers by Friston and colleagues describing variations of the generalized linear modeling (GLM) approach to filtering fMRI BOLD signal [1–4], which models an observed BOLD signal, $\mathbf{X}$, as

$$\mathbf{X} = \mathbf{H}\cdot\mathbf{w}_S + \mathbf{D}\cdot\mathbf{w}_C + \eta \tag{1}$$

where $\mathbf{H}$ is a matrix that comprised a set of explanatory functions (i.e., kernel vectors); $\mathbf{w}_S$ is the linear weight vector of these explanatory kernel vectors; $\mathbf{D}$ is a matrix comprised of a set of temporally structured confounding processes, $\mathbf{w}_c$; and, the final term, η is white Gaussian noise. This framework is the bedrock on which nearly all subsequent fMRI analyses are based. Early debate in fMRI modeling focused on the magnitude and role of autocorrelated noise processes, as well as the structure and applicability of the HRF [5,6]. Numerous papers debated the pros and cons of voxel-wise temporal smoothing and filtering [5,7,8].

The consensus from this early work, formed at the turn of century, is that temporal smoothing, in general, damages the

---

* Corresponding author at: Department of Computer Science, University of Arkansas at Little Rock, 2801 S. University Ave., Little Rock, AR 72204, USA. Tel.: +1 501 569 8143; fax: +1 501 569 8144.

E-mail address: kabush@ualr.edu (K. Bush).

http://dx.doi.org/10.1016/j.mri.2014.03.002
0730-725X/

underlying signal in fMRI except in cases of appropriate experimental design combined with band-pass filtering [8]. Also, autocorrelation (i.e., cardiac and respiratory influences) is the dominant source of noise and should always be modeled in concert with Gaussian white noise (i.e., thermal and quantum effects) [6,9,10].

Early work on modeling and smoothing BOLD signals was predicated on the use of SPM [11] as a means of identifying statistically significant task-related activations. As the GLM and SPM approaches matured, and sophisticated understanding of the HRF functions became available, researchers migrated the focus of fMRI BOLD signal modeling efforts toward identification of causal relationships in neural processing, particularly the general problem of capturing the underlying temporal distribution of neural events. This is exemplified in dynamic causal modeling, where deconvolution of the observed signal into neural estimates is the basis of forming causal inferences. Indeed, researchers [12–14] have subsequently argued for the necessity of deconvolution of the BOLD signal into its mediating neural events (either implicitly or explicitly) in order to improve inferences about neural activity.

Deconvolution is an inversion of the observed BOLD signal into a temporal distribution of individual neural events. This inversion process has been well studied, and numerous recent algorithmic approaches to this problem may be found in the literature [15–18]. The majority of deconvolution algorithms (excluding Bayesian filtering approaches [16]) assume the GLM form of the BOLD signal [15,18] in which matrix $\mathbf{H}$ (see Eq. (1)) is a convolution (i.e., Toeplitz) matrix formed from temporal offsets of the canonical HRF. Thus, the solution of this system yields a maximum likelihood estimation of the underlying true BOLD signal given quasi-physiological constraints. What makes deconvolution a unique problem is the allowable form of the weight matrix, $\mathbf{w}_S$, such that neural activations exhibit positive values and temporal structure, e.g., clustering and sparsity [18].

Convolution of the appropriate HRF with the correct weight vector, $\mathbf{w}_S$, will generate the true BOLD signal, i.e., a BOLD signal containing only neural estimates combined with their neurovascular consequences detected by BOLD. This unique feature of deconvolution (i.e., removing BOLD fluctuations that are not feasibly caused by neural influences) allows for a potentially powerful filtering approach: deconvolution of an observed BOLD signal into its neural estimates followed by reconvolution with an HRF estimate to produce a BOLD signal that is filtered of non-neural sources of influence (confounds) and retains the true signal of interest (neural processes). Thus, reconvolution of the deconvolved signal weights (termed the encoding), under certain assumptions of the system's physiology, constitutes an optimal filter, which we term deconvolution filtering.

The goal of this work is to examine and understand the viability of deconvolution filtering in improving analysis of real-world fMRI data. To achieve this, we perform a three-part analysis. In the first part, we simulate resting state fMRI BOLD of a single neural node using a number of known confounds: normalization scale-effects, downsampling effects, autocorrelated noise processes, and Gaussian white noise processes. The use of a simulation allows us to precisely estimate the root mean squared error (RMSE) and correlation between the filtered signals and the true, unconfounded BOLD signals. In the second part of the analysis, we vary additional confounds such as HRF mispecification, noise process autocorrelation, Gaussian white noise processes, and connection topology in a three-node neural system; the use of a simulation in this case allows us to precisely estimate the true functional connectivity graph of the three-node system, thereby facilitating quantitative assessment of both the absolute and relative roles of the various confounds, both within a single filtering approach and across differing filtering approaches: band-pass linear filtering, optimized band-pass linear filtering, as well as deconvolution filtering using both a GLM [15] and a nonlinear method [17] that strictly enforces positive neural representations. We deem analysis of functional connectivity estimation error critical in assessing how BOLD signal filtering impacts real-world, commonly used analysis. In the third part, we examine statistical properties of functional activation results of real-world fMRI BOLD data as additional motivation for the use of deconvolution filtering in practice.

We structure the manuscript into four parts. First, we describe a computational model of the BOLD signal's generation and observation for unifying performance results across filtering approaches. Second, we describe the deconvolution filtering framework, as well as two different deconvolution algorithms that will be used inside the framework during experimentation; we also briefly review two common band-pass filtering approaches and an optimized band-pass filtering approach that we use as benchmark comparisons on which to judge the performance of deconvolution filtering. Third, we describe and execute a set of experiments on both simulated and real BOLD that examine the performance of deconvolution filtering under numerous signal confounds. Finally, we combine the experimental results into a single assessment of the practical application of deconvolution filtering, and we offer guidance on fruitful directions of future research.

2. Methods

Our methods are composed of (1) a comprehensive parametric computational model of an fMRI BOLD signal's generation, (2) a novel algorithm for filtering a BOLD signal via its deconvolved representation, (3) band-pass filters based on Fourier analysis, and (4) an optimal Fourier-based band-pass filter; we describe these components below.

2.1. A generative model of fMRI BOLD data

Following prior simulation models [19,20], we construct a model of resting state neural activity using standard assumptions about the nature of the observed BOLD signal. We model fMRI BOLD data in four distinct processes: (1) a functional network generation process; (2) a neural generation process that captures temporal structure of neural events induced either by external, unmodeled processes or communication between brain regions; (3) a theoretical BOLD generation process that maps neural events onto an ideal set of BOLD signals (using either the canonical HRF or the balloon model); and, (4) an observation process that maps theoretically ideal BOLD signals onto low-frequency, noise-corrupted BOLD observations that represent real-world data. Each of these processes is described in detail below.

2.1.1. Functional network generation

We describe the relationships in space and time between V distinct brain regions as a functional network, $\mathcal{F} = \{\mathbf{C}, \mathbf{L}, \rho\}$, comprised of a connectivity model, $\mathbf{C}$, a communication lag model, $\mathbf{L}$, and an external activity model, ρ. The connectivity model, $\mathbf{C}$, is a matrix of conditional probabilities, $\mathbf{C}(i, j) \in [0, 1]$, $i, j \in 1,\ldots, V$, determining the probability that a neural event in brain region j will be induced by a neural event in brain region i. The lag model, $\mathbf{L}$, is a real-valued matrix, $\mathbf{L}(i, j) \in [0, L_{\max}]$, $i, j \in 1,\ldots, V$ dictating the time delay (in seconds) required for brain region i to influence region j. The bound $L_{\max}$ is the anatomically constrained maximum time required for a neural signal to traverse the brain. The external activity model, ρ, is a real-valued vector, $\rho(i) \in [0, 1]$, $i \in 1,\ldots, V$, which determines the probability of neural activity intrinsically generated by brain region i. Note, this intrinsic activity is intended to