

Comparison of Frequency-domain Statistics for fMRI Voxel Selection in Retinotopic Mapping

By

Niki Ziai

B.S. (University of California - Berkeley) 2001

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Biomedical Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Michael F. Insana

George R. Mangun

Thomas C. Ferree

Committee in Charge

2005

TABLE OF CONTENTS

1	Introduction	1
2	Functional MRI overview	4
2.1	Basic concepts	4
2.1.1	MRI physics	4
2.1.2	Functional MRI	6
2.2	Experimental design	10
2.3	Estimating experimental effect	13
2.3.1	Preprocessing	13
2.3.2	Response detection methods	15
2.4	Inference	17
2.4.1	Temporal autocorrelation due to noise	18
2.4.2	Theoretical vs. empirical distributions	19
3	Signal detection statistics for periodic designs	21
3.1	Correlation statistic: Co	21
3.2	Fundamental Power Quotient: FPQ	23
3.3	Magnitude Squared Coherence: MSC	26

4	Receiver operating characteristic curve analysis	30
4.1	Method overview	30
4.2	Results	34
4.2.1	Simulated Noise and Simulated Activation Datasets	34
4.2.2	Real Noise and Simulated Activation Dataset	41
5	Statistical validation	43
5.1	Method overview	44
5.2	Results	46
6	Retinotopic Mapping	50
6.1	Method overview	50
6.2	Results	55
7	Discussion	58
A	Hemodynamic modeling of superimposed signal	61
A.1	Balloon model	61
A.2	Neurovascular coupling	62
A.3	MRI physics	63
A.4	Model parameters	63

B	Simulating autocorrelative noise process	64
C	Relationship of C_0, FPQ, and MSC to theoretical distributions	64

LIST OF FIGURES AND PHOTOGRAPHS

1	Typical fMRI experimental paradigms	11
2	Regions of superimposed activation indicated in red in sample images of (a) fully-simulated and (b) partially simulated datasets	33
3	ROC curves for white noise dataset: simulation #1	35
4	ROC curves for phase-coherent noise dataset: simulation #2	37
5	Power spectra of representative voxel time series from: (a) Fig. 4c simulated dataset, (b) Fig. 7 real dataset, (c) simulated f^{-1} -structured noise spectra (see Appendix B for method) dataset	39
6	ROC curves for doubled TR and cycles/scan datasets: simulation #3	40
7	ROC curve for real null dataset: (a) full curve, (b) region of interest .	41
8	Two ways of interpreting bias using ROC curves	43
9	Actual false positive rates (FPR) for $\alpha = 0.05$ as a function of fre- quency: (a) simulated dataset, (b) real-null dataset	46
10	False-positive characteristic plot	48
11	Retinotopy visual stimuli: (a) polar angle wedge rotates either clock- wise or counter-clockwise, (b) eccentricity ring expands or contracts .	51
12	Structural MRI image processing: (a) raw image, (b) segmented white matter, (c) reconstructed cortical mesh	52

13	Polar angle phase data overlaid on: (a) coplanar anatomical scans, (b) MPRAGE, (c) reconstructed cortical mesh (d) flattened mesh	52
14	Determining borders of visual areas: (a) unthresholded flattened polar angle phase map of right hemisphere, (b) color legend relating phase to location of stimulus in visual field	54
15	Flattened polar angle phase maps of right hemisphere computed at fundamental frequency and thresholded for $p = 0.05$ using empirically determined distributions of each statistic	57

LIST OF TABLES

1	ROC curve mean in range FPR 0-0.1: simulation #1	35
2	ROC curve mean in range FPR 0-0.1: simulation #2	38
3	ROC curve mean in range FPR 0-0.1: simulation #3	40
4	ROC curve mean in range FPR 0-0.1: real noise simulation	41
5	Actual false positive rates	47
6	MRI pulse sequence parameters for scans used in retinotopy	53

1. INTRODUCTION

This project aims to improve the sensitivity of functional magnetic resonance imaging (fMRI) of human visual cortex. This goal will be pursued in a specific context, that of the retinotopic mapping of visual cortex using fMRI, but should also be of utility in analysis of imaging data outside this application.

Retinotopic mapping is the identification of regions of visual cortex that contain complete, spatially contiguous maps of the visual world, known as cortical *visual areas*. Research in non-human primates that recorded single neuron activity and investigated the cellular microanatomy of the visual cortex revealed that the visual cortex of the primate has a successive series of these visual areas that can be recognized by their complete retinotopic map of the visual world.¹ These visual areas are organized in a hierarchy, and in humans it is possible to employ fMRI to non-invasively obtain retinotopic maps of visual cortex. Determining the borders of different visual areas allows us to examine the response properties of those areas to a wide variety of visual stimuli under different task circumstances. Such research is important for understanding the functional organization of the human brain and is essential for understanding how disease and damage affect the basic organization and function of the brain.

Analyses of functional MRI data falls into two broad categories: those that assume an explicit hemodynamic response function (fMRI response to instantaneous neural impulse) and those that do not. As is the case with many fMRI studies and

almost always with retinotopy, periodic stimulation designs are used. These allow the option of frequency domain analysis and thus avoid the use of explicit hemodynamic models. In an effort to minimize assumptions, we pursue frequency domain analysis of the BOLD time series within mrVista, an open-source tool designed for retinotopic mapping.

While mrVista implements a signal detection statistic, it is a rather basic statistical test and a number of other statistical methods have been reported in the literature that are purported to be more sensitive. We have studied and implemented a number of these in mrVista. Three frequency domain statistics will be discussed: mrVista's Co (a correlation statistic - as defined by Engel et al., 1997),² Fourier power quotient (FPQ - as defined by Bullmore et al. 1996),³ and magnitude squared coherence (MSC). The first two statistics have been defined and used in previous fMRI studies, but to our knowledge this paper explores for the first time the application of MSC to fMRI. These statistics allow us to search for responses to periodic stimuli that occur at the frequency of stimulation (and/or its harmonics) in fMRI voxel time series. The objectives of this research are to compare the performance of Co , FPQ , and MSC using simulations and receiver operating characteristic (ROC) curve analysis, explore their validity under various assumptions, and examine the results of their application to real retinotopy data.

Implementing these various tests in mrVista accomplishes several things. First, it has provided a venue in which to learn all the techniques of frequency-domain fMRI analysis and retinotopic mapping in detail. Second, it allows us to examine for the

first time whether these techniques improve retinotopic mapping. Third, it makes these other statistical techniques readily available to researchers in this field.

This last point is significant, because while research in statistical analysis of fMRI data moves forward rapidly, most researchers use one of a small number of packages to perform their analyses and the introduction of these novel statistical techniques into these packages is relatively slow. By implementing novel algorithms within the framework of an existing software package, we extend the immediate capabilities of research in the Center for Mind and Brain which advances ongoing NIH- and NSF-funded research at UC Davis.

2. FUNCTIONAL MRI OVERVIEW

While the localization of brain function has long been known from effects of traumatic brain injury and in the mid 20th century from cortical stimulation during neurosurgery, the imaging of brain function has only existed since the middle to late 20th century with the advent of positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). Since the 1990s, fMRI has come to dominate the brain mapping field due to its non-invasiveness, lack of radiation exposure, and relatively wide availability.

A brief review of the underlying physics, experimental design, and data analysis of the functional MRI is presented to provide the background for the research analysis techniques compared in this thesis.

2.1. Basic concepts

2.1.1. MRI physics

MRI is based on the phenomenon of nuclear magnetic resonance (NMR). During a scanning session, the subject being imaged is placed in a uniform magnetic field of great strength (typically 1.5 or 3 Tesla) where the magnetic dipole moments of atomic nuclei with non-zero spin numbers within the tissue align either parallel or anti-parallel to the main field. In human brain imaging it is the hydrogen nuclei that are of interest. Their great abundance in tissue allows the small percentage of their magnetic dipoles which prefer one orientation to the other (lower energy state:

parallel; 1/million at equilibrium) to produce a detectable change in the magnetic field.

Before any manipulation however, the local net difference in magnetization, M_o , is orders of magnitude less than B_o and thus not directly observable. In order to generate a measurable signal, a radio frequency pulse is applied at the resonance frequency of the nuclei of interest (governed by the Larmor equation: $\nu_o = \gamma * B_o$: ν_o = frequency of precession; B_o = main magnetic field strength; γ = gyromagnetic ratio) causing the dipoles to tip or *nutate* into a plane perpendicular to the axial field. This process of nutation takes the nuclei from their aligned low-energy state into a temporary non-aligned high-energy state and occurs with the operation of three orthogonal gradients coils. Referred to as the slice selection, phase encoding, and frequency encoding gradients, these gradients produced by the three coils allow for spatially localized differences in magnetic field and thus spatially specific excitation. The strength of these gradients can be defined in various combinations giving the user the capacity to prescribe almost any image slice orientation desired.

Because of the intrinsic quantum mechanical spin properties of nuclei, when in the aligned or non-aligned states, the dipoles of the hydrogen protons precess around the axis of the main field (B_o) at their characteristic Larmor frequency. The precessing dipoles of the non-aligned nuclei, made phase coherent by the radio frequency pulse, will *relax* back to their low-energy state. As they relax, the coherent precession of many spins produces a changing magnetic field that induces a measurable current in a nearby coil. The detected signal is known as *free induction decay*.

Images with contrast that emphasize specific tissue properties of interest can be obtained by manipulating the MR signal's level of sensitivity to different tissue properties through the pulse sequence parameters. The primary types of contrast available are known as proton density, T1, and T2—each emphasizing a different property of the three dimensional local magnetization vector, M , that is generating the signal for each voxel of the image slice. Proton-density-weighted images are sensitive to the equilibrium longitudinal magnetization, M_o . T1-weighted images are sensitive to the relaxation time constant of the longitudinal component of M . And T2-weighted images are sensitive to the relaxation time constant for the transverse component of M . In fMRI, image contrast is based on a decay constant known as T2*. T2* is an apparent T2 time constant in MR signals generated by pulse sequences sensitive to magnetic field inhomogeneity. This is appropriate for acquisition of functional information of the brain as described more below.

2.1.2. Functional MRI

Functional MRI in the broadest sense is the use of MRI technology for measuring local physiology rather than its common use for structural anatomy. Most commonly, fMRI is the modality for detecting hemodynamic changes resulting from neural activation using the Blood Oxygenation Level Dependent (BOLD) contrast effect discovered in 1990 by Ogawa and colleagues.⁴⁻⁶

In the range of all functional brain mapping techniques, fMRI has a unique combination of features that make it particularly amenable to the study of sensory, motor,

and cognitive activities in the human brain. Because of the need both to identify these dynamic activities and to localize them, functional brain imaging methods are often compared and contrasted in terms of their temporal and spatial resolution. Most exhibit a distinct trade off. Electrophysiological (EEG) and magnetoencephalography (MEG) measures, for example, map transient electrical and magnetic dipoles resulting from neuronal depolarization with great temporal resolution (10-100 msec); but as the measurements are typically recorded at the scalp, EEG and MEG have relatively poor spatial resolution (several mm to cm). Single-unit electrophysiology refines spatial localization to the neuronal level (less than 1 mm), but requires placing electrodes in the cortex and thus is neither safe nor practical. Functional MRI, as well as PET and optical imaging methods (such as near infrared spectroscopy), measure neural activity indirectly through the slower time-scale dynamics (on the order of seconds) of hemodynamic changes resulting from metabolism and accompanying increases in blood flow. With slightly higher temporal resolution than fMRI and PET, optical imaging suffers in spatial resolution unless the cortical surface is exposed, which then like single-unit recordings poses a significant safety issue for the subject and practicality issue for the experimenter. And even then, optical imaging measures are restricted to the cortical surface. Functional MRI and PET, on the other hand, achieve relatively high spatial resolution (1-10 mm) within deep structures of the brain. While PET is moderately invasive, requiring injection of a radioactive tracer, fMRI has the unique distinction of achieving better spatial and temporal resolution than PET and higher spatial resolution than EEG/MEG through an entirely non-invasive procedure

that involves minimal risk for the subject. Though lacking the ability of EEG and MEG to measure the temporal dynamics of neural activity, fMRI is a practical and powerful tool for localization brain function with time scales that are still useful and continually improving.

BOLD fMRI relies on the changes in the concentration of the endogenous paramagnetic contrast agent hemoglobin. When neurons fire in a region of the brain, the metabolic demand from neurotransmitter recycling reduces the blood oxygenation in the region, which in turn alters the magnetic susceptibility of the blood. This produces after a lag of a few seconds a slight but measurable increase in the T2*-weighted MR signal that in turn also lasts a few seconds (peaks at about 6 sec). Initially, the regional increase in oxidative metabolism results in a transient increase in the concentration of paramagnetic deoxyhemoglobin in the blood, increasing the local non-uniformity of the field near vessels, dephasing local spins, and decreasing the MR signal but usually to undetectable levels. Shortly after (approximately 1 sec), an over-compensatory increase in cerebral blood flow (CBF) to the region quickly supplies excessive amounts of oxygen, decreasing the deoxyhemoglobin concentration, reestablishing the field uniformity, and consequently increasing the T2-weighted MR signal. These measurable blood oxygenation dependent changes are on the order of 1 to 5 percent greater than baseline using fMRI and constitute what is referred to as the BOLD effect. The BOLD fMRI temporal impulse response function to a brief stimulus, known as the hemodynamic response function (HRF), is consequently delayed and blurred in time.

The most common pulse sequence employed to measure the BOLD effect (and the one employed for the data collected in this research) is known as the gradient-recalled-echo (GRE) echo-planar-imaging (EPI) sequence. Gradient-recalled-echo sequences are particularly sensitive to the neurally-induced local inhomogeneities in the magnetic field that are responsible for the BOLD effect (as described in the previous paragraph). When compared to other contrast sequences, GRE sequences produce larger activation-induced signal changes by a factor of 2 to 4. The detected free induction decay signal (FID) of this GRE sequences is characterized by the decay rate, $R2^* = 1/T2^*$, or the characteristic decay time, $T2^*$. During the time course of the scan, the decay rate decreases slightly during periods of activation, increasing the measured signal. The positive and negative variability across time in signal due to these activity-induced changes in the decay rate of the FID defines the BOLD effect.

The fast MR imaging techniques in the EPI aspect of the pulse sequence allow a series of images of the brain to be acquired in rapid succession. With EPI it is possible to acquire BOLD image data from a single slice of the brain in under 0.1 sec (the entire brain in about 2 sec) and with spatial resolution greater than 3x3x3mm. Rather than acquiring a single image line (in k space) after the preparation phase of the pulse sequence as often is otherwise the case with MRI, the entire 2D MR image is acquired between the radio frequency pulses of the GRE-EPI sequence. Many adjacent 2D image slices are acquired to span at minimum the regions of interest in the brain. To distinguish activity from rest, a series of functional data volumes of the brain are acquired in succession while the subject is stimulated or performs a

task that probes the function(s) of interest at only certain points in time throughout the scan. These successive acquisitions form a time series of measured data points for each individual resolved element (voxel) in the volume that can subsequently be processed off-line and tested for significance.

2.2. Experimental design

A wide variety of experimental designs exist for probing and measuring neural activity during a fMRI scan. All these designs follow the general scheme of supplying the brain with particular stimuli in the hope that regions involved in processing that stimuli respond with an increase (or possibly a decrease) in metabolic activity that is measurable with the BOLD signal. Due to the subtlety of the BOLD effect, the statistical quality of a single image is not adequate to detect the signal of interest. To alleviate this, usually several scans are collected under the same experimental conditions and averaged together. Even throughout the duration of a single scan, however, the main goal of all the various stimulation paradigms is to repeat the experimental condition as many times as possible to maximize the signal to noise ratio of each voxel time course (here the effective SNR being the ratio of stimulus-related variability to non-stimulus-related variability across a time course). Some are chosen for simplicity of design and analysis and others for maximal efficiency, where efficiency is defined as the ability to accurately estimate stimulus-related signal in the allotted scan time. The hypothesis being tested, the assumptions one is willing to make about the nature of neural activity, as well as certain constraints (e.g. minimizing

scan time) dictate the most appropriate choice for a given study. In general, they fall into three broad categories: block, event-related, and periodic designs.

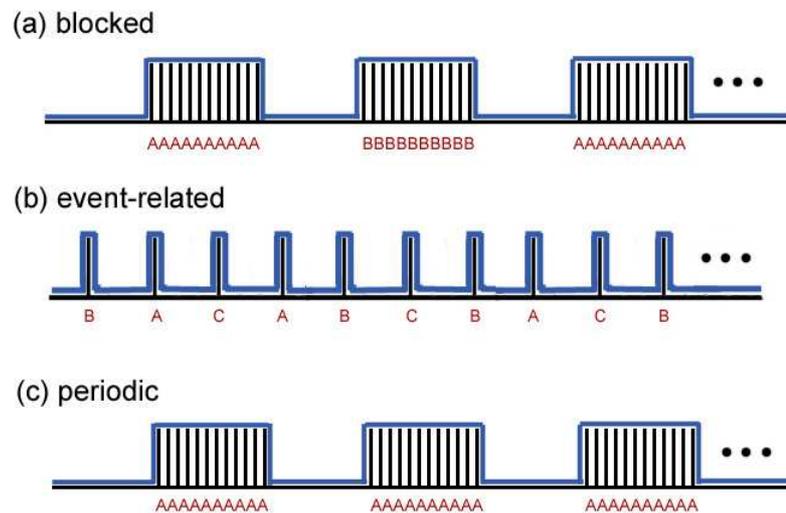


Figure 1. Typical fMRI experimental paradigms

The most basic of these is known as the block design in which trials for stimuli of a particular condition (for example A or B in Fig. 1(a)) are grouped together and presented in blocks of time and ‘off’ for others. The premise for such a design is that neural activity is additive such that subtraction of activity during the ‘off’ (or baseline) period from activity during the ‘on’ period leaves only stimulus related activity. Both stimulus ‘on’ and ‘off’ periods last about 16-40 seconds allowing time for the generation of sustained hemodynamic response and its subsequent decay. The resulting data can be analyzed for stimulus-related activity several ways including t -tests on mean signal differences between conditions,⁷ cross-correlation with the stimulus time-course vectors,⁸ multiple linear regression⁹ to name a few. If the block

design is periodic, Fourier analysis is also an option with distinct advantages. Linear systems theory—in which the fMRI signal is assumed to vary linearly with the local average neuronal response averaged over time plus noise^{10,11}—is the framework for many of these methods.

The most common analysis technique used today is based on the general linear model (GLM) and like some of the other above listed methods involves the use of an assumed hemodynamic response function. The hemodynamic response function in the cortex has tremendous variability both between and within subjects,¹² nevertheless, theoretical and estimated models for the response function are often used.

Another common design is the event-related design. Rather than grouping multiple trials of a particular condition in blocks, in event-related designs different types of stimuli (A, B, and C, in Fig. 1(b)) are presented in arbitrary sequences of single trials.^{13–15} This design has the advantage of eliminating potential confounds such as habituation, anticipation, or other strategy effects. While not as efficient as the block design in accurately estimating the stimulus-related activity, variations in the timing and sequence of trials have been shown to provide significant improvement in efficiency.¹⁶ Analysis of event-related data can be done with selective averaging¹⁷ or linear modeling¹⁶ both of which require an assumed or estimated HRF.

This thesis focuses on periodic designs. A periodic design is a special case of blocked designs. In a periodic design, stimulus trials of the same condition (A in Fig. 1(c)) are presented in blocks of equivalent duration that repeat periodically

throughout the run. The advantage of periodic designs is that data can be analyzed entirely in the frequency domain where the variance attributable to stimulus response will be restricted to a few discrete frequencies, namely the fundamental and its harmonics. Like previous techniques, analysis in the frequency domain also assumes that the fMRI signal varies linearly with the local average neuronal response. Unlike previous techniques, however, frequency domain analysis avoids explicit assumptions about the hemodynamic model. This is an advantage because such models are not generally available for each voxel in each subject's dataset.

2.3. Estimating experimental effect

Identification of stimulus-related activity relies upon hypothesis testing in which the null hypothesis, H_0 , is the absence of any stimulus-related activity, and the alternate hypothesis, H_1 , its presence. Classification of data as satisfying either H_0 or H_1 usually occurs at the individual voxel level and involves prerequisite processing steps. After the functional MRI dataset has been acquired and preprocessed, the response detection method of choice is applied to estimate the magnitude of stimulus-related activity. These results are subsequently tested for statistical significance.

2.3.1. Preprocessing

Functional MRI datasets are subjected to a variety of preprocessing steps before searching for response to the stimuli. These steps are both spatial and temporal. The most essential ones are summarized below.

Spatially one needs to correct for subject motion during the scan. Although subjects are instructed to stay still and stabilized as much as possible in the scanner with the help of cushioning, temple supports, and sometimes even bite bars, inevitably there is some movement at the very minimum from normal physiological processes. Because subsequent processing assumes that data from a particular voxel represents a particular part of the brain, if the degree of movement exceeds the voxel size, failing to correct for motion will result in displaced activations in the fMRI time-course and contribute to the variance of the statistical test. Motion correction algorithms vary but have been shown to be comparable in their performance.¹⁸ In mrVista and in general, the 6 parameter (3 rotation, 3 translation) estimates of an affine rigid body transformation are determined from and applied to the reconstructed image data.

Temporally, fMRI datasets are preprocessed to correct for slice timing differences. Although subsequent analysis assumes so, slices in a functional volume are not acquired at the same point in time. Often the 2D data in the 3D functional volume is acquired in an interleaved manner first acquiring odd numbered slices and then even (or the reverse) to minimize 'inflow' effects between adjacent slices. This can cause data from adjacent slices to be as far as $TR/2$ apart in time (where TR is the scan repeat time). Slice timing algorithms correct for this by shifting the phase of the voxel time series appropriately. Although this is a common preprocessing step in most fMRI analysis software packages, mrVista did not employ it and we did not implement it for our analysis. Because each statistic is computed on each voxel independently, the relevant timing between slices is irrelevant to voxel selection. In drawing borders for

retinotopic maps (see section 6.1), spatial phase coherency between adjacent voxels becomes important. We recognize that this is an issue, but because this study focuses on the signal detection aspect of retinotopy, we did not focus on it. However, any subsequent analysis which requires or involves averaging over voxels will be sensitive.

2.3.2. Response detection methods

After the data has been run through the appropriate preprocessing routines, it is then ready for analysis. Response detection techniques fall into two broad categories: temporal correlation using the general linear model and Fourier analysis.

By far the most common method as mentioned earlier is the general linear model (GLM), otherwise known as multiple regression.^{9,19} Based on the principles of linear time invariant systems, for any of the stimulus designs described above, the GLM can be used to model the signal time course in each voxel using linear combinations of multiple basis functions called *regressors*. In matrix form, the GLM is as follows:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{r} \quad (1)$$

where \mathbf{y} is the $N \times 1$ column vector representing a voxel time course (N being the number of time points); \mathbf{X} is the $N \times M$ matrix (known as the *design matrix*) containing the regressors of interest (M being the number of $N \times 1$ regressors); \mathbf{b} is the $M \times 1$ column vector of model parameters to be estimated; and \mathbf{r} is the $N \times 1$ matrix of residual error terms assumed to be independent and normally distributed with mean 0 and variance σ^2 .

The regressors in design matrix \mathbf{X} include those that describe non-stimulus-related trends and those that describe stimulus-related trends. Typical non-stimulus-related regressors include baseline and linear drift. Stimulus-related regressors, $g(t)$, are defined by convolving the hemodynamic response function (HRF) with the time course for each stimulation condition, i :

$$g_i(t) = \sum_{u=-\infty}^{\infty} h(u)s_i(t-u) \quad (2)$$

where $h(t)$ is the HRF (typically assumed, sometimes estimated) and $s(t)$ the stimulation time course.

The model is fit to each voxel's time course giving minimum least squares estimates of the parameters \mathbf{b} . These estimated parameters reflect the degree of correlation of each regressor. The resulting parameter for each $g_i(t)$ is then usually divided by its standard deviation to form a t -statistic from which significance levels can be derived.

When the stimulus paradigm is periodic, Fourier analysis may be applied. Using the discrete Fourier transform (DFT),

$$Y(f) = \frac{1}{N} \sum_{t=1}^N y(t)e^{-i2\pi tf} \quad (3)$$

we can compute the signal power at all frequencies 0 to Nyquist and look for significant peaks at the fundamental (paradigm) frequency and its harmonics. Letting the stimulus design consist of q cycles of the stimulation-rest period the DFT frequencies of interest will then be those corresponding to the set $[q, 2q, 3q, \dots, N/2]$. This approach avoids the use of an assumed HRF (although still assuming the brain's

functional response can be approximated by an linear time invariant system) and is the standard for retinotopy. To estimate the magnitude of stimulus-related activity there are numerous ways to process the spectral information. The three signal detection statistics, Co , FPQ , and MSC , each do this differently and are the subject of investigation in this thesis.

2.4. Inference

While response detection is a critical step in the analysis process, it is not complete without the ability to make accurate inferences about the measured effect. Drawing conclusions as to whether voxel time-series have passed (H_1) or failed (H_o) the hypothesis test requires estimates of the significance of the measured effect. Accurate estimation of the significance in turn requires that the effect be detected without assumptions that cause bias and that the distribution from which significance thresholds are determined be legitimate representations of the chosen statistic's distribution. If the first condition is not met (i.e., assumptions of the signal detection statistic being used are violated), the magnitude of the signal can be overestimated, biasing the detection statistic in sensitivity. If the second condition is not met (i.e. unrepresentative distribution), the false positive rate can be either under- or over-estimated, biasing the signal detection statistic in specificity. These two conditions are related in ROC curves. To avoid either inflated or overly conservative significance levels, both require robust approximation.

2.4.1. Temporal autocorrelation due to noise

A well-documented confound in estimating the significance of effect in fMRI time-series is that of temporal autocorrelation due to noise. Zarahn et al., for example, collected BOLD fMRI data on human subjects who did not perform any time-locked experimental behavior. They found that when statistical analysis of these ‘real-noise’ datasets involved assumptions of independence, the null-hypothesis was rejected at a significantly higher rate than the expected significance level, α .²⁰ Purdon and Weisskoff likewise reported disparity between actual false positive rates and assumed significance levels in synthesized (using a autoregressive plus white noise model) null datasets.²¹

In fact, numerous studies including the above two, have explored the nature of the autocorrelation in fMRI noise time series and methods of accounting for it. Zarahn et al. found that the noise power spectra was described well by a f^{-1} model exhibiting disproportionate power at low frequencies. Investigating the sources of the autocorrelative noise, they found that the f^{-1} was characteristic of both their human noise datasets as well as water phantom datasets indicating that it was not strictly a physiological phenomenon. They likewise found that the f^{-1} component could not be completely explained by the first-order effects of motion, equipment present during scanning, or the convolution of neural activity with hemodynamic response functions as used commonly in GLM analysis techniques. In attempting to account for autocorrelation they found that temporal smoothing with low pass filters in addition to

Worsley and Friston’s modified GLM technique resulted in false positive rates close to the theoretically expected significance level. The technique suggested by Friston et al. is to *coloring* or *shaping* the intrinsic autocorrelation with a larger known autocorrelation the majority of which can be removed with high-pass filtering.^{9,19} Others, including Bullmore et al., have suggested procedures for estimating the autocorrelation in order to prewhiten the data in a GLM framework^{3,22–24} and resampling techniques that allow for inference that is unbiased by temporal autocorrelation in the noise.²⁵ Although best methods for accounting for temporal correlation in the noise continue to be explored, there is clearly wide agreement that it must be addressed to achieve unbiased inference, particularly in the context of low-frequency paradigms in which the consequences of autocorrelation are more pronounced.

2.4.2. Theoretical vs. empirical distributions

While assumptions of independence in the noise time series in the face of temporal autocorrelation can bias actual false positive rates away from the assumed α , assumed distributions from which significance thresholds are chosen can potentially have the same effect. Often times, signal detection statistics are shown to vary *like* a well-known statistic’s distribution such as a the F -, Student’s t -, or the complementary error function, *erfc*. However, they actually do so only when they meet certain assumptions—the most common being that they were computed on data whose noise processes were independent and normally distributed. If the data does not meet the necessary assumptions, the theoretical distributions of the statistics will not match

the true distributions, and significance levels will be biased either above or below the expected α . The only way to control for this is to empirically determine the statistical distribution using a real noise dataset. This controls for bias in inference but does not correct the effect of autocorrelation in the noise.

3. SIGNAL DETECTION STATISTICS FOR PERIODIC DESIGNS

In the analysis of functional MRI data (periodic or not), there are a large variety of signal detection statistics that can and have been employed in the detection of regional brain responses to various behavioral stimuli and tasks. The intention of this investigation was not to do an exhaustive comparison of them all, but to begin to with a common measure already used in retinotopy, consider the most appropriate existing alternative, and explore the relative performance of a never before used but directly relevant and potentially advantageous technique. In this light, the three statistics of focus in this thesis: *Co*, *FPQ*, and *MSC* are introduced below.

3.1. Correlation statistic: *Co*

The first statistic we examine is known as *Co* and was defined by Engel et al. in their 1997 paper on retinotopic mapping.² Having been introduced in one of the founding papers of retinotopy and implemented in the mrVista toolset, *Co* has become the default statistic for many researchers. As shown in the Engel et al. paper, *Co* was derived from the standard definition for temporal correlation of two column vectors, where the vectors represent a harmonic function at the stimulus frequency and the data time series for a particular voxel. The derivation (explicitly shown in their paper) results in the following definition:

$$Co(f_s) = \frac{|Y(f_s)|}{\sqrt{\sum_f |Y(f)|^2}} \quad (4)$$

where $Y(f)$ is the Fourier transform of $y(t)$, the average (over repeated scans) measured BOLD time series in a single voxel; f_s denotes the stimulation (paradigm) frequency; and the sum in the denominator is over all frequencies from zero to Nyquist. Co therefore ranges from zero to one, equaling zero when there is no power at the paradigm frequency and equaling one when there is only power at the paradigm frequency. Assuming a purely harmonic stimulus, which is always the case with retinotopy, Co tells you how much signal amplitude you have at the stimulus frequency channel compared to other places in the spectrum.

In light of periodic stimulus designs, the goal of Co is to classify fMRI voxels as positive (satisfying H_1) when the power at the stimulus frequency dominates the voxel's power spectrum. While there is logic to this approach, Co has the drawback of potentially failing to recognize truly active voxels when there is a significant amount of power at frequencies other than f_s . Power that exists at other frequencies may be signal or noise and will increase with increased signal bandwidth. Co cannot tell the difference between the sources of this power. A significant proportion of power contained in the first few harmonics of f_s , for example, will be signal but inappropriately contribute to reducing Co by being included in its denominator. Power from scanner and physiological noise, likewise all frequencies irrelevant to the activity of interest, will exist and contribute to reducing Co . While scanner noise is typically white, often in fMRI studies, one will find additional power at low frequencies due to temporal autocorrelation of the sort discussed above and cardiac and respiratory frequencies, all of which contribute detrimentally to signal detection with Co . And

when the data is acquired at higher sampling rates, the increased bandwidth of the signal will lead to power at previously non-existent frequencies that again get added to Co 's denominator and reduce Co 's value. The potential saving grace of Co is that it is computed on the averaged time-series. Because artifactual noise with random phase across scans will cancel upon averaging, the noise power in the denominator will decrease leading to an increase in Co 's magnitude.

3.2. Fundamental Power Quotient: FPQ

The next statistic we study is called the fundamental power quotient (FPQ), defined by Bullmore and colleagues.³ As indicated by its name, the FPQ statistic, like Co , addresses the issue of power at the paradigm or *fundamental* frequency and thus also is intended for periodic designs. Likewise, FPQ is computed on the averaged data. Unlike Co however, the analysis that leads to FPQ is performed in the time domain, rather than the frequency domain. The approach of Bullmore et al. is to implement pseudogeneralized least square (PGLS, also known as the Cochrane-Orcutt transformation), an iterative form of ordinary least squares (OLS) regression modeling. The aim of PGLS is to correct for autocorrelation in the noise time series, as manifested in the residuals of OLS, leaving the residuals of the second linear model close to white. To provide background for the FPQ statistic, a summary of Bullmore's PGLS technique is given below, followed by the definition of FPQ .

First, the BOLD time series of each voxel is fit, using OLS, to the following

regression model that accounts for linear and sinusoidal trends in the data:

$$\begin{aligned}
 y(t) = & \alpha + \beta t + \gamma \sin(\omega t) + \delta \cos(\omega t) + \gamma' \sin(2\omega t) \\
 & + \delta' \cos(2\omega t) + \gamma'' \sin(3\omega t) + \delta'' \cos(3\omega t) + \rho(t)
 \end{aligned} \tag{5}$$

where $y(t)$ is the average (across scans) signal for a single voxel, $t = 1, 2, 3, \dots, N$; ω is the fundamental frequency in radians per time point; α, β, γ 's, and δ 's are the model parameters to be estimated; and ρ is a residual term. Bullmore and colleagues suggested the above model based on some exploratory analysis of experimental BOLD fMRI data collected on a human subject undergoing periodic photic stimulation. Having averaged the time series of 156 voxels representing the occipital cortex, they inspected the plots of the averaged time series and corresponding periodogram and correlogram. From these plots they observed evidence over the course of the experiment of a slight negative linear trend, marked periodic or sinusoidal trend with the same fundamental frequency as the input function, and relatively modest peaks in the periodogram corresponding to the first and second harmonics of the fundamental frequency—all resulting in the suggested regressors in Eq. 5.

Equation 5 can also be represented in matrix notation as:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{r} \tag{6}$$

where \mathbf{y} is the N -dimensional column vector representing the T2*-weighted signal intensity (averaged over scans) of a single voxel at timepoints $t = 1$ to N , \mathbf{X} is the $N \times 8$ design matrix, \mathbf{b} is the 8-dimensional column vector of model parameters,

and \mathbf{r} is the N -dimensional column vector of residuals. In the case that the error terms are independent and normally distributed, OLS provides minimum variance unbiased estimates (MVUEs) of the model parameters. Bullmore et al. examined the residuals, using a partial autocorrelation function (PACF), and showed that they were autocorrelated to the first degree. To address this, they modeled the residual autocorrelation using a first-order autoregressive (AR) process:

$$\rho(t) = \zeta * \rho(t - 1) + \epsilon \quad (7)$$

The error terms of this AR model, ϵ , were shown to have no significant autocorrelation. This allows the subsequent use of the OLS estimated AR coefficient, ζ , to transform the original terms of the regression model as shown below (transformed terms are asterisked, *) and consequently incorporate the autoregressive nature of the data into the model:

$$\mathbf{y}^*(t) = \mathbf{y}(t) - \zeta \mathbf{y}(t - 1) \quad (8)$$

$$\mathbf{x}^*(t) = \mathbf{x}(t) - \zeta \mathbf{x}(t - 1) \quad (9)$$

Via corroboration with PACF plots of the PGLS residuals, the above modeling is then assumed to produce independent and normally distributed error terms. Having achieved MVUEs, the FPQ statistic is computed. The FPQ statistic is simply power at the paradigm frequency divided by its standard error, SE . It makes use of the $\tilde{\gamma}$ and $\tilde{\delta}$ model parameter estimates to do so, giving (under the assumption of independent

and normally distributed parameter estimates):

$$FPQ = \frac{(\tilde{\gamma})^2 + (\tilde{\delta})^2}{\sqrt{2(SE(\tilde{\gamma})^4 + SE(\tilde{\delta})^4)}} \quad (10)$$

While computed in the time domain, the linear regression that provides the model parameters for FPQ accomplishes something similar to a Fourier decomposition. The Fourier technique is a form of multiple linear regression. When computing the Fourier transform of a time domain signal as done for Co , the time series is decomposed into the sum of sine and cosine waves at all $N/2 + 1$ equally spaced frequencies. The regression model described above differs from a Fourier transform in that the time series is decomposed into only a restricted sum of sines and cosines, including only the fundamental and first two harmonics, and includes a linear trend and error term. Nevertheless, by using the coefficients of only the fundamental frequency, FPQ attempts to avoid being influenced by information from frequencies other than the fundamental. Assuming the 8 model parameters are adequate for modeling the signal, compared to Co , this may be advantageous.

For the purposes of our investigation, FPQ was incorporated into mrVista toolset so that it could be computed from the model parameters both before and after removal of autocorrelation in the residuals. This allows comparison of the effect of autocorrelation removal.

3.3. Magnitude Squared Coherence: MSC

Magnitude squared coherence is a frequency domain function used in numerous areas including for example system identification, signal-to-noise ratio (SNR) measurement,

and estimation of time delay. It can be applied to spatial or temporal data and is essentially a measure of the registration of two datasets at a particular frequency. For linear systems, coherence measures the proportion of response power attributable to a given stimulus, and thus can be thought of as a "signal to signal plus noise" ratio. In its most general form it can be represented as the square of the cross power spectral density of the two signals divided by their autospectral densities. Letting $Y(f)$ and $S(f)$ be the temporal Fourier transforms of two time series signals $y(t)$ and $s(t)$, the magnitude squared coherence, or MSC , is defined as:

$$MSC(f) = \frac{|G_{sy}(f)|^2}{G_{ss}(f)G_{yy}(f)} \quad (11)$$

where $G_{sy}(f) = S^*(f)Y(f)$ and $G_{yy}(f) = Y^*(f)Y(f)$. MSC is real-valued and like Co ranges from zero to one. Since MSC computed from a single pair of signals is equal to 1, MSC must be estimated from multiple samples. This estimation procedure shows that for signal detection applications where one is searching for a biological response to a perfectly periodic stimulus, $s(t)$, such as the functional MRI paradigms studied here, MSC depends only on the measured data. Derived from Eq. 11, this result follows as:

$$\begin{aligned} MSC(f) &= \frac{|G_{sy}(f)|^2}{G_{ss}(f)G_{yy}(f)} \\ &= \frac{|S^*(f)|^2|Y(f)|^2}{|S(f)|^2G_{yy}(f)} \\ &= \frac{|Y(f)|^2}{G_{yy}(f)} \\ &= \frac{|\frac{1}{M} \sum_{i=1}^M Y_i(f)|^2}{\frac{1}{M} \sum_{i=1}^M |Y_i(f)|^2} \\ &= \frac{|\sum_{i=1}^M Y_i(f)|^2}{M \sum_{i=1}^M |Y_i(f)|^2} \end{aligned} \quad (12)$$

where M is the total number of time series samples in the estimate.^{26,27} Coherence is shown here to equal the power of the mean response divided by the mean power of the subaverages. From this equation we see that, upon averaging, MSC reduces to a ratio of signal to signal plus noise, as mentioned above. Letting the Fourier transform of the measured time course for a single voxel $Y(f)$ be represented as a sum of response related signal and noise as $Y(f) = A(f) + N(f)$ where $A(f)$ is Fourier transform of the response to the stimulus and $N(f)$ is a combination of instrumental and physiological noise taken to be additive and zero-mean,

$$MSC(f) = \frac{A(f)^2}{A(f)^2 + N(f)^2} \quad (13)$$

Magnitude squared coherence has been used as a method of objective response detection of evoked potentials in electroencephalographic (EEG) data^{27,26} but to our knowledge has never before been used in fMRI analysis of brain activity. Given the definition above one can see the relevance of such a function to periodic fMRI analysis. Not only is MSC qualitatively appropriate because of its focused use of frequency information, but it has unique quantitative characteristics that save it from the difficulties faced by other statistics such as Co which take the entire frequency spectrum into account. MSC like FPQ uses information from only the stimulus frequency (although FPQ begins by fitting the signal to the stimulus frequency and its harmonics). In practice, however, the PGLS technique may not model the signal perfectly as does the Fourier transform (by Parseval's theorem) and thus the coefficients of the fundamental frequency that go into FPQ may not be as representative of frequency

information at the fundamental frequency as the Fourier terms used in *MSC*.

MSC was incorporated into the mrVista toolset using Eq.12 such that the Y_i' s represent voxel time-series segments derived from dividing the original time-series of each scan. Because the number of scans that can be acquired is limited (by expense and subject endurance) the process of dividing the time series for each scan into multiple segments was introduced as a method of maximizing the total number of time series samples, M in Eq.12. There exists a trade-off, however: because the length of the segments determines the frequency resolution of the power spectrum; for very short segments, the bin at the fundamental frequency encompasses more than just the response frequency. To minimize spectral leakage, voxel time-series were divided such that each resulting segment contained an integer number of stimulus cycles. Since in our retinotopy studies there were either 10 or 20 cycles per scan, various degrees of segmentation existed according to the above criterion. The implemented code computes *MSC* for all integer divisors. All the following reported results for *MSC* reflect computations based on the best performing divisor.

4. RECEIVER OPERATING CHARACTERISTIC CURVE ANALYSIS

To begin our comparison of these three statistics, we used the receiver operating characteristic (ROC) curve technique. ROC curves are a convenient method for comparing the performance of different detectors. In the 1950's, ROC curves were developed to distinguish radar signals from noise. Today, ROC curves are a popular tool in medical and imaging research. ROC curves were first introduced to the analysis of functional MRI data by Constable and colleagues in 1995 and have since been used extensively as a tool for objective comparison of fMRI signal detection strategies.²⁸⁻³¹ A summary of basic ROC methods for fMRI data and our results using the technique follows below.

4.1. Method overview

Using ROC curves, we can visually compare and quantify a detector's accuracy. An ROC curve represents accuracy in terms of true positive rate (TPR = sensitivity) vs false positive rate (FPR = 1 - specificity) at all possible threshold values of a detector. In order to create an ROC curve one must know the distributions of true signal and noise in the fMRI dataset. This is sometimes but not usually the case. When the true distributions are unknown, simulated data must be used. For ROC analysis from simulated data that is an accurate representation of its application to real data, the MRI images should contain noise and artifacts that are representative of the typical fMRI data under consideration. The best way to obtain this is using partially

simulated dataset where artificial activations are added to a real null experimental dataset, i.e. fMRI data collected on a human subject while no visual stimulation is presented and all other conditions are otherwise held constant. Since the noise in this type of data is measured and the true activation locations are known, the FPR (proportion of voxels incorrectly detected as active to all truly non-active voxels) and TPR (proportion of correctly detected active voxels to all voxels that contain added activation) values can be computed accurately. Computing the FPR and TPR at different thresholds for a particular detection statistic provides the points from which the ROC curve is fit. Because the ROC curve represents a system's complete sensitivity and specificity range, overlaying the ROC curves for different statistics visually displays their relative accuracy without the influence of a threshold value. Generally speaking, the further a curve lies above the one-to-one FPR-TPR line (chance detection line), the more it reflects a better detector.

Although signal detection statistics can be compared qualitatively by visually comparing their ROC curves, various methods also exist for quantifying ROC curve results in terms of a single metric. These methods differ in the proportion of the curve that they take into consideration. Common metrics include the area under curve, (A_z), or best operating point (point furthest from the diagonal) where detectors with greater A_z or best operating points furthest from the curve win.³⁰ Others advocate the use of only a small portion of the curve in calculating a metric, arguing that only lower FPR ranges are ever used in fMRI studies and thus the only relevant region from which ROC curve metrics should be computed.^{29,30} These groups use either the

area under the curve or the mean TPR within this restricted portion (often FPR 0 to 0.1). Because conclusions for fMRI data are only ever drawn from highly significant results (typically p-values < 0.05), in our use of this tool to compare Co , FPQ , and MSC statistics, we chose to use the later metric computing the mean of the ROC curve in the range 0 to 0.1.

Our computation of ROC curves proceeded on two different dataset types: one containing both artificial noise and artificial activations and the other containing only artificial activations. The first was based on fully simulated data to facilitate characterization of the performance of the three statistics under varying noise and paradigm conditions. In all these simulations, signal was placed in roughly 10% of the voxels (approximate percent of typically active voxels in an image; see Fig. 2(a)), in each of four scans subsequently averaged together and was created by summing sine waves at the typical retinotopy stimulus cycling frequency (0.0265Hz) and its first two harmonics. The noise for each of these simulations, on the other hand, was systematically modulated in its magnitude, phase coherence across trials (not voxels), and frequency range to determine each statistic's sensitivity not only to overall differences in signal-to-noise ratio but to changes in noise structure as well. Finally, since SNR will vary with imaging rate and the number of volumes (timepoints) per scan, in another subset of these simulations, we also tried doubling our typical retinotopy imaging rate (TR) and number of cycles per scan (initially set as TR=1.57 and number of timepoints = 240; see Chapter 6) to see what if anything the effect would be.

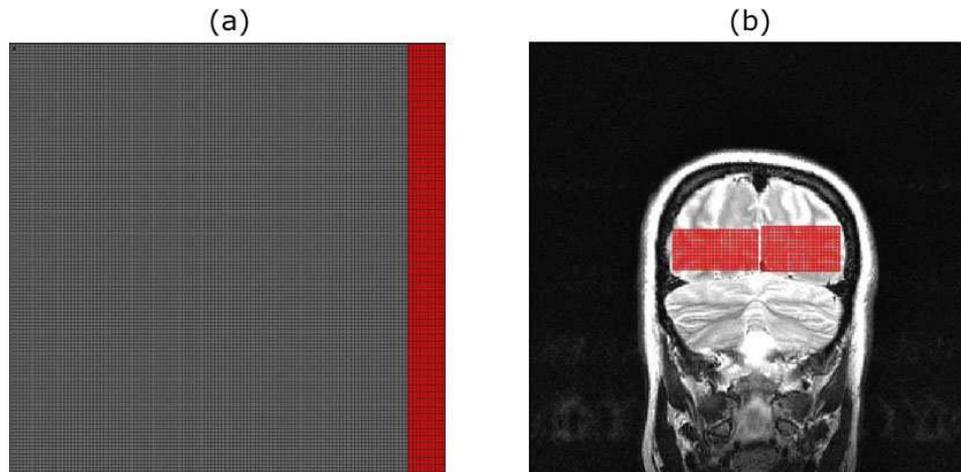


Figure 2. Regions of superimposed activation indicated in red in sample images of (a) fully-simulated and (b) partially simulated datasets

The second dataset type was based on partially simulated data as described above where real null data was collected under experimental conditions perfectly identical to our typical retinotopy sessions less the visual stimulation, and periodic activation was superimposed on brain voxels (seen overlaid on a high resolution anatomical T2-weighted image acquired coplanar to our functional scans in Fig. 2(b)). Specifically, four 7-minute EPI scans (see Table 6 for parameters) were collected with a human volunteer who performed the typical retinotopy task of passive fixation where this time, however, the screen remained isoluminant gray for the duration of each scan. Since target activity is always in the brain and brain voxels differ from non-brain voxels in that they contain not only scanner noise, but also physiological noise, our superimposed activations were restricted to brain voxels. In doing so we ensured that any detected activations were only classified as such after being subjected to

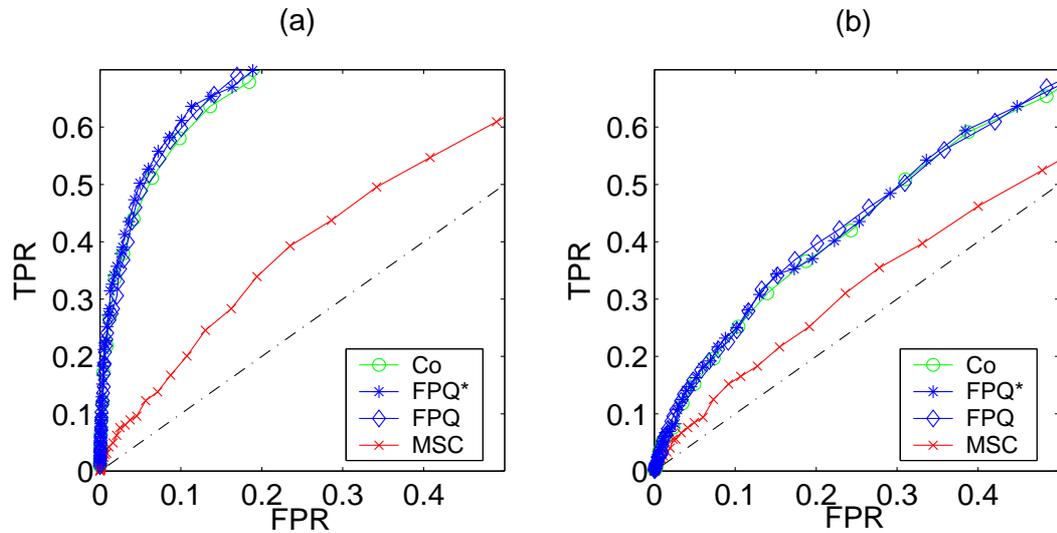
the same noise challenges as real data. Likewise, the superimposed signal in this case was made to closely resemble actual hemodynamic response through the use of the Balloon Model proposed by Buxton³² with neurovascular coupling by Friston³³ (Appendix A).

4.2. Results

4.2.1. Simulated Noise and Simulated Activation Datasets

We began our experimentation with fully simulated datasets by superimposing ten cycles ($f_s = 0.0265Hz$) of the sine wave signal on 10% of 4096 (64x64 image) randomly generated (normally distributed, mean=0, variance=1) voxel timeseries (240 timepoints). The amplitude of the sine wave was varied as a function of the percentage of noise standard deviation (σ). As seen in the ROC curves of Fig. 3, both Co and FPQ perform better than MSC at low signal-to-noise ratios: (a) signal amplitude = 0.1σ [SNR=0.10] (b) signal amplitude = 0.05σ [SNR=0.05] where FPQ^* is FPQ computed before accounting for autocorrelation in the residuals of the regression model. Since noise was white by definition, there was no residual autocorrelation in the data and FPQ^* and FPQ lines predictably overlap. As signal-to-noise ratio is decreased, comparing Fig. 3(a) to Fig. 3(b), the ROC curves for all 3 statistics approach the chance detection line. At even minimal SNRs, MSC fails to perform better than the other statistics. The results for our chosen ROC curve metric—mean TPR in FPR range (0, 0.1)—shown in Table 1 prove only the same. Co and FPQ beat MSC by a factor of nearly 2 for SNR=0.05 and a factor of more than 4 for

SNR=0.10.

**Figure 3.** ROC curves for white noise dataset: simulation #1**Table 1.** ROC curve mean in range FPR 0-0.1: simulation #1

	SNR	Co	FPQ^*	FPQ	$MSCdF3$
Fig. 3(a)	0.10	0.4314	0.4579	0.4378	0.1063
Fig. 3(b)	0.05	0.1421	0.1486	0.1479	0.0869

Co and FPQ clearly have an advantage over MSC in situations of purely random noise. This can be explained (at least partly) by the fact that Co and FPQ are computed on voxel time-series that have already been averaged over scans. In the case of random noise, averaging reduces the power at all frequencies other than the paradigm frequency (and its harmonics) [since response to the stimulus will be phase locked across scans] significantly boosting the relative power at the paradigm

frequency and thus also both the sensitivity and specificity of Co and FPQ which are both in essence measures of just that: relative power at a paradigm frequency.

MSC , on the other hand, while also a subject of averaging (albeit a different method), only uses information from the paradigm frequency. This may provide an advantage over Co and FPQ in situations where a significant amount of noise remains at *other* frequencies after averaging over scans, but otherwise may require more scans or lengthier time-series (increasing M in Eq. 12) for MSC to have sufficient power distinguishing signal from signal+noise at the stimulus frequency to outperform FPQ and Co . For noise to remain in the spectrum after averaging, the noise processes would have to be coherent in their phase across scans. In reality it is unlikely that the periodic noise processes of cardiac and respiratory artifact are time locked to the stimulus cycling period and even less likely that their phases will be coherent across scans. We were, nevertheless, interested in exploring such cases in order to empirically gauge the range of MSC 's theoretical strengths.

To test the effect of noise coherency across scans on these statistics, we created another simulated dataset in which the same sinusoidal signal as before was superimposed on time-series that contained phase-coherent noise. These noisy time-series were generated by superimposing on the randomly generated time-series (as above) a series of sine waves at all frequencies from twice the paradigm frequency ($2f_s = 0.0531Hz$) up to Nyquist ($0.3185Hz$), all with phase equal to π . We initially began with this broad coverage of the spectrum and set the amplitude of the phase coherent noise equal to that of the signal (0.1σ) to take maximum advantage

of *MSC*'s insensitivity to power at non-paradigm frequencies as accorded by theory. We likewise restricted the phase-coherent noise to the 10% of voxels that contained signal, as phase-coherent noise if analogous to anything in real data is most likely to be analogous to physiological noise and thus would be restricted to brain voxels. Then, across a series of these datasets we incrementally reduced the spectral range of phase-coherent noise by $2f_s$ frequencies from the low end of the range. Figure 4 shows the progression of ROC curves as the frequency range of phase-coherent noise is reduced and Table 2 shows the results of our ROC metric.

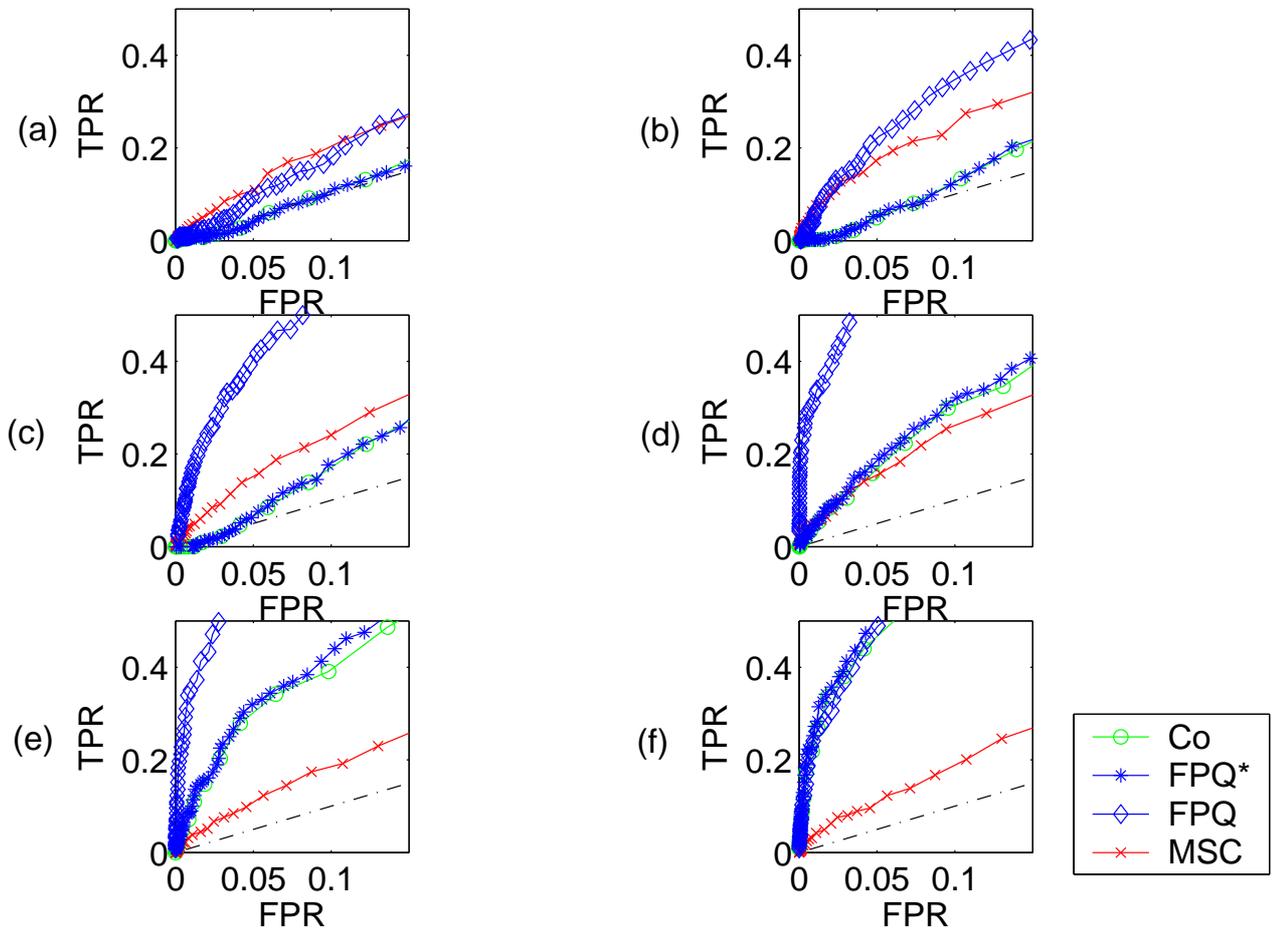


Figure 4. ROC curves for phase-coherent noise dataset: simulation #2

Table 2. ROC curve mean in range FPR 0-0.1: simulation #2

	Freq Range	Co	FPQ^*	FPQ	MSC
Fig. 4(a)	$2f_s+$	0.0466	0.0451	0.0878	0.1152
Fig. 4(b)	$4f_s+$	0.0509	0.0574	0.2022	0.1597
Fig. 4(c)	$6f_s+$	0.0699	0.0729	0.3638	0.1422
Fig. 4(d)	$8f_s+$	0.1655	0.1762	0.5133	0.1505
Fig. 4(e)	$10f_s+$	0.2645	0.2786	0.5391	0.1062
Fig. 4(f)	none	0.4314	0.4579	0.4378	0.1063
	variance	0.0233	0.0263	0.0321	$5.597e^{-4}$

The ROC curves of Figure 4 demonstrate some important features of the 3 statistics. Primarily, as the range of phase-coherent noise is reduced, MSC remains rather stable while the other statistics shift significantly. The variances of our ROC metric for these statistics (see Table 2) across the datasets also attests to this, being two orders of magnitude smaller for MSC than the others. Resistance to non-paradigm frequency noise is clearly the strength of MSC . Next, it is important to note that while MSC is robust in this way, it performs better than FPQ and Co only when the range of phase-coherent noise is substantial (Fig. 4(a-c)). Finally, we recognize that FPQ^* and FPQ are dissociated in the presence of coherent noise but FPQ^* is consistently virtually indistinguishable from Co . Because FPQ differs from FPQ^* , the PGLS algorithm for residual autocorrelation removal is influential here. Its behavior in this particular simulation, however, should not be interpreted in the context of

real data as its noise structure (Fig. 5(a)) is different from the f^{-1} spectral behavior characteristic of real data (Fig. 5(b)).

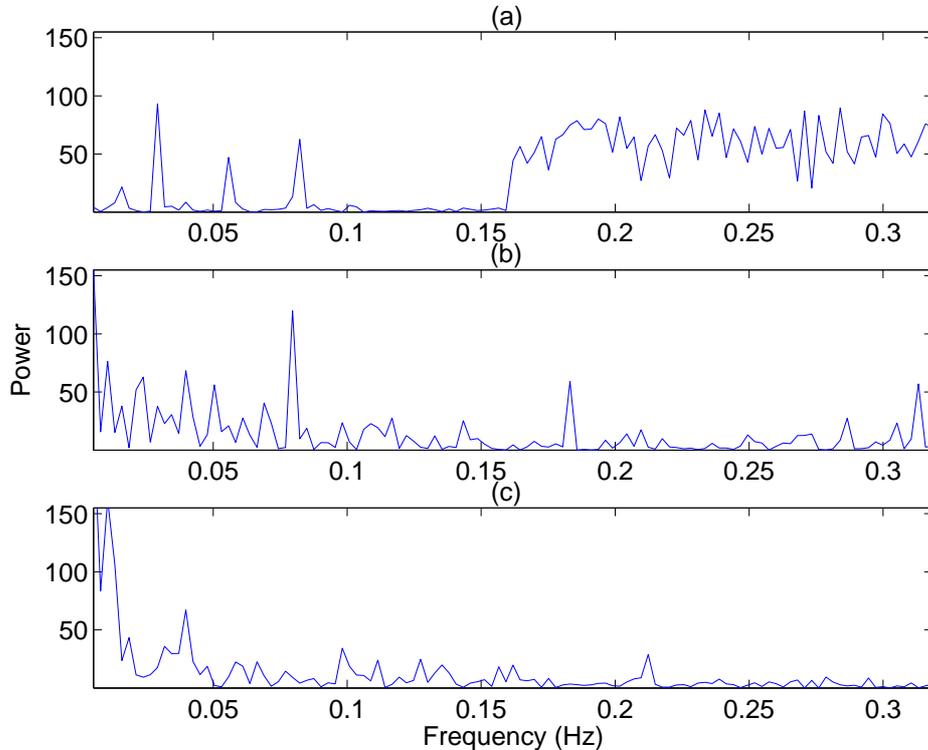


Figure 5. Power spectra of representative voxel time series from: (a) Fig. 4c simulated dataset, (b) Fig. 7 real dataset, (c) simulated f^{-1} -structured noise spectra (see Appendix B for method) dataset

Having evaluated the sensitivity of the statistics to noise magnitude and coherency, our final datasets of the fully simulated type were constructed to test performance variability in response to changing imaging rate and time-series length. Although our typical retinotopy EPI scans make use of a TR of 1.57 seconds, typical TRs in fMRI studies are often about twice as long. Time-series lengths can likewise vary. As a comparison to our previous simulations where TR=1.57 and number of timepoints =

240, ROC curves resulting for datasets in which TR and cycles/scan were doubled (independently) are shown in Fig. 6 with corresponding metrics in Table 3.

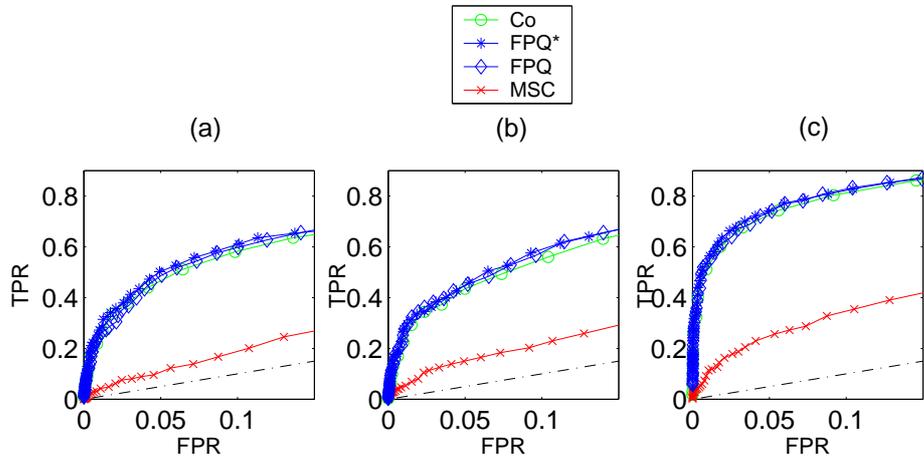


Figure 6. ROC curves for doubled TR and cycles/scan datasets: simulation #3

Considering the results, we see that doubling the TR has had no noticeably significant effect. Doubling the number of cycles per scan has slightly improved the performance of all the statistics, but the relational pattern of performance is the same: Co and $FPQs$ are collinear and superior to MSC .

Table 3. ROC curve mean in range FPR 0-0.1: simulation #3

	TR	cycles/scan	Co	FPQ^*	FPQ	MSC
Fig. 6(a)	1.57	10	0.4314	0.4579	0.4378	0.1063
Fig. 6(b)	3.00	10	0.4023	0.4263	0.4221	0.1396
Fig. 6(c)	1.57	20	0.6852	0.7054	0.6976	0.2289

4.2.2. Real Noise and Simulated Activation Dataset

Having explored the behavior of Co , FPQ , and MSC by synthesizing and manipulating characteristics of the noise in fully artificial datasets, we proceeded to test the ROC performance of datasets containing real noise. The results for ROC analysis of this partially simulated real null dataset (methods described in chapter introduction) are shown in Fig. 7 and Table 4.

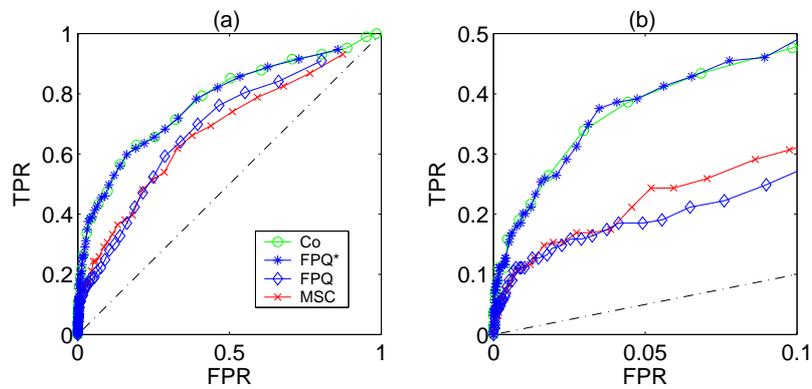


Figure 7. ROC curve for real null dataset: (a) full curve, (b) region of interest

Various features of the curve in Fig. 7 deserve attention. Initially, one will notice that Co and FPQ^* are collinear and perform better than MSC and FPQ in the region of interest. The collinearity of Co and FPQ^* matches findings from previous simulations (simulations#1-3) as does the dissociation of FPQ^* and FPQ (simulation#2). From this dissociation we can conclude that the OLS residuals must not

Table 4. ROC curve mean in range FPR 0-0.1: real noise simulation

	Co	FPQ^*	FPQ	MSC
Fig.7	0.3629	0.3649	0.1834	0.2103

be independent and normally distributed otherwise FPQ would coincide with FPQ^* as in simulation#1 (Fig. 3) where only normally distributed random noise was used. Likewise, since Co and FPQ^* are once again coincidental and separate from FPQ , we may surmise that Co is not accounting for temporal autocorrelation due to noise if that is indeed the only difference between FPQ and FPQ^* , i.e. all PGLS does for FPQ is prewhiten the data.

Finally, it is interesting and important to note that FPQ 's reduced performance compared to FPQ^* is on the order of that of MSC 's, MSC slightly outperforming FPQ according to our metric (Table 4). If, as their definitions suggest, Co and FPQ^* are deficient in their capacity to correct for residual autocorrelation and FPQ appropriately addresses it, FPQ should be taken as the valid statistic and Co as invalid. With this in mind, taking the fact that MSC was proven to be stable (i.e. show less variability) to severe manipulations in noise structure (simulation#2: Fig. 4 and Table 2) and the fact that MSC is commensurate with FPQ in this real null dataset, there exists the possible conclusion that MSC is robust to autocorrelation by nature, functioning without the need for explicitly accounting for temporal autocorrelation in the noise.

5. STATISTICAL VALIDATION

While ROC curves provide a comparison of signal detection performance, before drawing any conclusions about the superiority of a statistic from them, one must be confident of having met the assumptions of the response detection method as well as any assumptions about the statistic's distribution when thresholding. Both these factors influence the validity of inferences drawn from the statistics (see section 2.4 for review). If either factor is not addressed appropriately, the obtained p -values and consequently inferred significance levels will be biased.

A test may be considered *valid* if it is accurate in the statistical inferences it makes. As previously discussed, the assumption of independent noise time series (manifested in the successive terms in the error time series in linear modeling) can bias results when temporal autocorrelation is not accounted for during response detection and when theoretical distributions are used for thresholding. Whether it is left

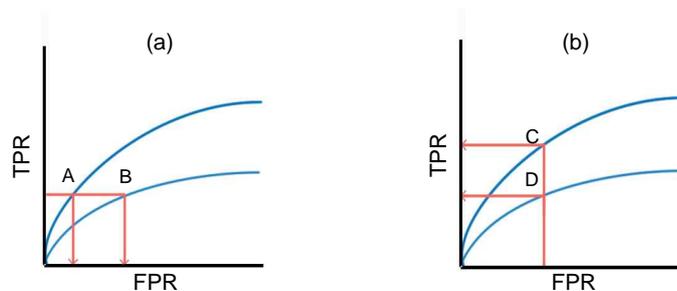


Figure 8. Two ways of interpreting bias using ROC curves

unaccounted for in the response detection method that produces a signal detection statistic or if it is assumed in the use of a theoretical distribution for thresholding,

the result of such an assumption would manifest itself as the falsely superior ROC curve seen in *both* images of Fig. 8. The bias exemplified similarly in both Fig. 8(a) and Fig. 8(b) can be understood in two ways. Looking at Fig. 8(a), we can see that for a given TPR, a biased curve (A) will give reduced false positive rates (B). Then, considering Fig. 8(b), we see that for a given FPR, a biased curve (C) will give greater than actual true positive rates (D).

We explored the existence of the two causes of this bias in usage of Co , FPQ , and MSC as described below.

5.1. Method overview

Because our real null dataset simulations resulted in ROC curves (Fig. 7) that reflected the condition displayed in Fig. 8, the goal of our first validation test was to see how if at all the residual autocorrelation structure that was dissociating FPQ^* from FPQ was biasing actual significance levels away from assumed ones. In other words, we wanted to explore the effect of assuming independent noise on bias in statistical significance. Using null datasets (both simulated and real) we tabulated how often the value of each statistic passed an assumed significance threshold of $p = 0.05$. Because of the use of a null dataset, this qualification should in theory happen no more than 5% of the time. If its rate of occurrence were to exceed 5%, we would have evidence of bias. To hold variability in thresholding technique constant and at the same time use a technique that reveals sensitivity to noise-autocorrelation as a function of frequency, each statistic's $p = 0.05$ threshold was determined from its

distribution across frequencies. Specifically, for each of 1000 simulated-null voxels and 1700 real-null voxels separately, Co , FPQ^* , FPQ , and MSC values were computed for all available frequencies and ordered in a histogram to determine the cutoff value for the top 5%. For each voxel and at a particular frequency then, the statistic was qualified as either having passed or not passed its cutoff threshold, the number of times it passed over the total number of tests (i.e. 1000 or 1700) constituting its actual FPR.

Simulated null datasets were generated to have inherent autocorrelation. This was achieved using an algorithm proposed by Rangarajan and Ding in which the spectral density of a white noise process is made to scale with frequency, f , as the power law f^{-1} (see Appendix B).³⁴

To test the effect of chosen distribution on the validity of inference, we again used a real null dataset but this time determined thresholds using theoretical distributions. We compared how well resulting FPRs in the null dataset matched a range of assumed α 's in the form of an 'actual FPR vs. assumed alpha' (or "False-positive characteristic" as termed by Purdon and Weisskoff) plot.²¹ The theoretical distributions were chosen based on previous associations to the spectral statistics discussed here and included the t -statistic and $erfc$ (complementary error function) for Co ,^{8,35} Chi -squared for FPQ ,³ and an F -statistic for MSC ²⁶ (exact relationships to of statistics to these distributions made explicit in Appendix C). Actual FPRs were computed separately for ten 64x64 image slices of the null dataset and averaged to produce the values plotted in the figure.

5.2. Results

Figure 9 shows the results of our test for bias resulting from the effect detection assumption of independence in the noise time series. Because we set the threshold

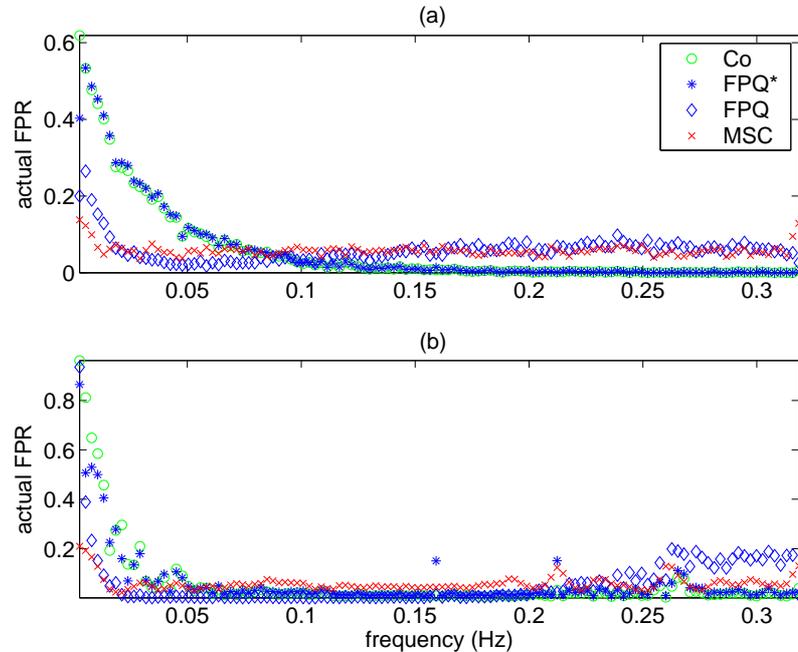


Figure 9. Actual false positive rates (FPR) for $\alpha = 0.05$ as a function of frequency: (a) simulated dataset, (b) real-null dataset

at $p = 0.05$, valid conservative statistics will have false positive detection rates (FPRs) equal to or less than 0.05. As we see in both the simulated- and real-null datasets, at low frequencies, actual FPRs for all four statistics do exceed 0.05. In the range of 0-0.05Hz, however, *MSC* and *FPQ* have values closer to the assumed α . Indeed, in the simulated dataset (Fig. 9(a)), *Co* and *FPQ** seem to scale like the f^{-1} power law structure we had incorporated into the data, while *MSC* and *FPQ* are virtually flat except at the front tail. While this is not as evident in the real-null dataset, the trend

Table 5. Actual false positive rates

	Co	FPQ^*	FPQ	MSC
Fig. 9(a)				
$f = 0.0265$	0.234	0.239	0.054	0.057
$f = 0.0530$	0.106	0.110	0.023	0.058
Fig. 9(b)				
$f = 0.0265$	0.134	0.134	0.005	0.047
$f = 0.0530$	0.029	0.034	0.005	0.030

is the same. It could be that the magnitude of noise autocorrelation was minimal in this particular real dataset, nevertheless we recognize that at our low retinotopy paradigm frequency of 0.0265Hz, Co and FPQ^* significances are still inflated by a factor of nearly 3 while that of MSC and FPQ are below the assumed (Table 5). Based on the qualitative results of these graphs and quantitative values summarized in Table 5, we would conclude that the greater the amount of temporal autocorrelation in the noise and the lower the paradigm frequency, the more caution needs to be taken in the choice of statistic. MSC and FPQ would not just be the more conservative choices, but more likely the valid ones.

Figure 10 shows the results of our test for bias as a result of thresholding based on theoretical distributions. $Erfc$ and t -stat distributions commonly used with Co have produced a positive bias, inflating the significance while Chi -squared and F -stat distributions used with FPQ and MSC have produced a negative bias, underestimating

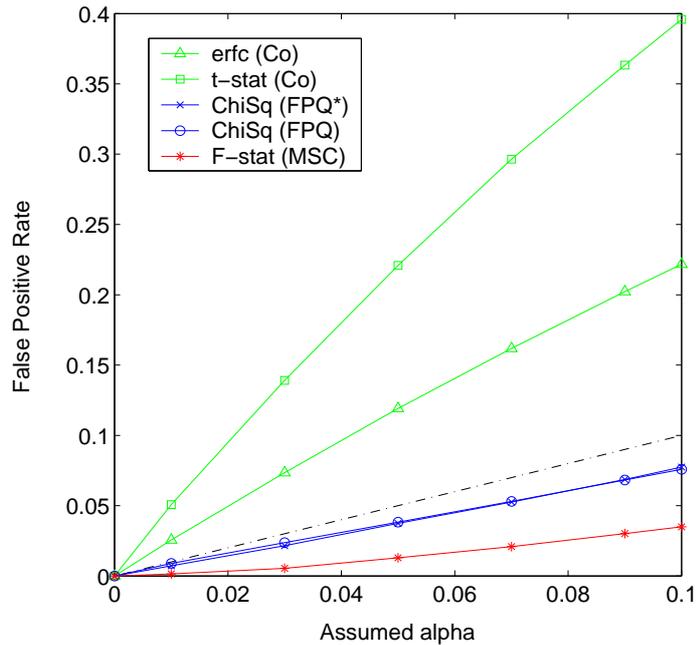


Figure 10. False-positive characteristic plot

the significance. This plot argues in favor of determining thresholds empirically. Because a real-null dataset is required, empirical thresholding is inefficient in both cost and time; however, the benefit is virtually perfect knowledge of the null distribution. In the case that empirical thresholds are unavailable, results of this plot advocate the use of the *Chi*-squared distribution with *FPQ*, being slightly more conservative than the assumed. And although *FPQ** and *FPQ* are coincident, considering the results of Fig. 9, *FPQ* should still be used over *FPQ** to account for bias due to temporal autocorrelation in the noise.

Our explorations of bias due to assumptions in effect detection and in thresholding has shown that both do exist and are important concerns. In studies using low paradigm frequencies, such as retinotopy, when independence of error terms is incor-

rectly assumed in signal detection, significance levels for Co and FPQ^* are inflated above the assumed α . On the other hand FPQ , which accounts for temporal autocorrelation, and MSC , which may be robust to it, remain more consistently below the assumed α and thus are the more conservative and valid options. Likewise, when the ideal empirical distributions are unavailable, thresholding based on theoretical distributions can bias false positive rates either above or below the assumed α . If theoretical distributions must be used, the statistics with the more conservative false positive rates are preferred to ensure validity in this account.

Having investigated the general performance and validity of the Co , FPQ , and MSC statistics, we proceeded to examine their results in the context of retinotopic mapping.

6. RETINOTOPIC MAPPING

Retinotopic maps that relay orderly representations of visual space from the retina to the brain are a universal feature of the visual system of vertebrates. With the advent of functional MRI and the aid of cortical reconstruction and flattening technology, it has been possible to non-invasively determine the retinotopic organization of the human visual cortex.^{36,37,2} Borders of visual areas can subsequently be derived therefrom. Analysis of fMRI data for retinotopy differs from most fMRI studies in that the paradigm is a phase-encoded periodic one, and the data must be visualized on 2D representations of the cortex. The following gives a brief summary of the processing steps involved and carried out in our analysis with the mrVista toolset as well as comparison of signal detection results for *Co*, *FPQ*, and *MSC* on a real dataset.

6.1. Method overview

Neurons of retinotopic visual cortex are responsive to only limited areas of the visual field, known as their receptive fields. The stimuli used for retinotopic mapping exploits this property to map the polar angle and eccentricity of points in the visual field to the cortex. Black and white checkerboard stimuli spanning as much of the visual field as possible (approximately 17 degrees of visual angle in our case) and reversing at 8Hz, the optimal frequency for visual neuron excitation,³⁸ of both wedge (Fig. 11(a)) and ring (Fig. 11(b)) formations are rotated and expanded/contracted respectively about the fixation point in the center of the visual field during separate EPI scans.

While the subject passively fixates on the cross during these scans, the stimuli produce

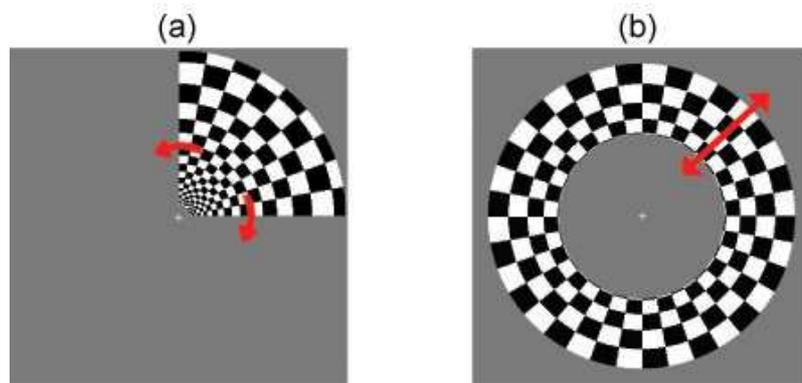


Figure 11. Retinotopy visual stimuli: (a) polar angle wedge rotates either clockwise or counter-clockwise, (b) eccentricity ring expands or contracts

waves of activation across the cortex as neighboring neurons with receptive fields in the direction of the probed visual field coordinate (i.e., polar angle or eccentricity) fire. The stimuli cycle across the visual field at a slow rate (0.0265Hz for wedges and 0.0531Hz for rings) activating the cortex at the same frequency. Since the cortex is activated at the same frequency of the cycling stimulus, we expect stimulus-related response in fMRI time series to occur predominantly at the frequency of stimulation. The frequency of stimulation thus is a filtering criteria for response detection, and the phase information from each voxel in the functional volume correlates the location of the stimuli to particular regions of the cortex. In this sense, the stimuli and brain activity are *phase-encoded*. Using the discrete Fourier transform this phase information is easily extracted for each voxel. Only certain regions of the brain are responsive to the stimuli, however, and that is where signal detection statistics come into play. For each voxel in the volume a signal detection statistic is computed and

thresholded.

To visualize retinotopic data in the most convenient manner, the functional phase information is overlaid on flattened representations of the cortex generated from high-resolution T1-weighted anatomical MRI data from the same subject. First the high-

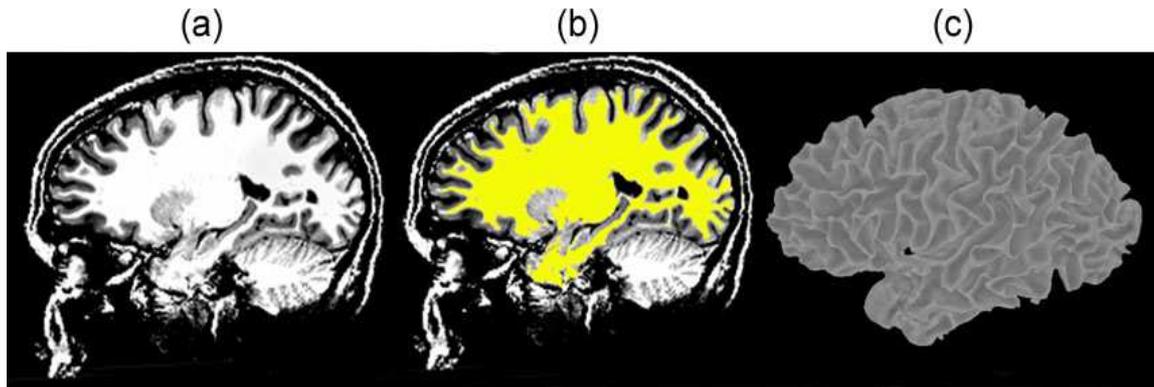


Figure 12. Structural MRI image processing: (a) raw image, (b) segmented white matter, (c) reconstructed cortical mesh

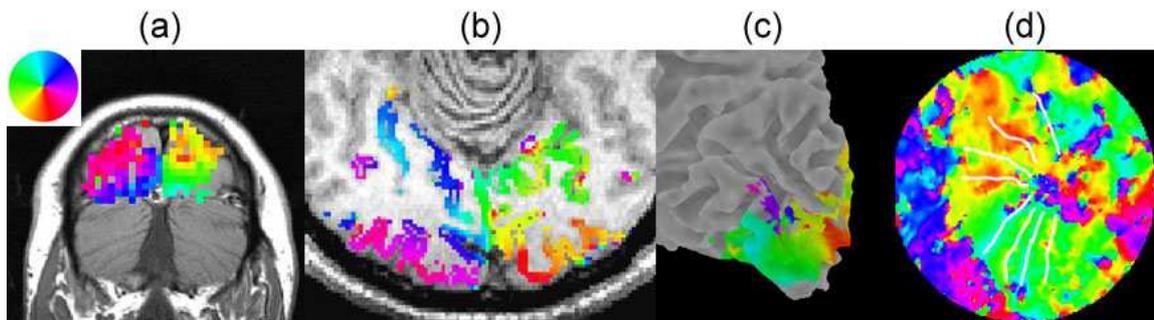


Figure 13. Polar angle phase data overlaid on: (a) coplanar anatomical scans, (b) MPRAGE, (c) reconstructed cortical mesh (d) flattened mesh

resolution anatomical scan (MPRAGE, Fig. 12(a)) is segmented along the gray-white matter border (Fig. 12(b)) and reconstructed into a 3D cortical mesh (Fig. 12(c)).

This 3D mesh is flattened in the occipital pole region. Next high-resolution *coplanar* T2-weighted 2D anatomical scans acquired with the same orientation and slice thickness as the functional volumes are aligned to the 3D MPRAGE coregistering the datasets. Finally, using the registration parameters, the functional phase-maps can be visualized directly on the coplanar anatomical scans (Fig. 13(a)), the high-resolution 3D anatomical (Fig. 13(b)), the folded cortical mesh (Fig. 13(c)), or the flat mesh (Fig. 13(d)). It is the flattened views that allow us to clearly see retinotopic patterns on the cortical surface and determine areal borders (indicated by white lines in Fig. 13(d)). MRI scanning parameters for the various pulse sequences used are listed in Table 6.

Table 6. MRI pulse sequence parameters for scans used in retinotopy

Description	Matrix	FOV (mm)	TR (ms)	TE (ms)	FA (degrees)
MPRAGE	256x256	256	2500	4.8	7
T2-TSE	256x256	220	4000	25	90
GRE-EPI	64x64	220	1570	25	90

Definitions of retinotopic visual areas are determined by identifying complete and spatially contiguous regions of the phase map representations of the visual field across the cortex. Borders between the areas are determined by identifying contiguous regions of reversal in the polar angle phase gradients (Fig. 14). By examining each hemisphere’s flattened functional phase maps, one will notice reversals in phase occurring at the phase values representing stimulation of the vertical meridian and the

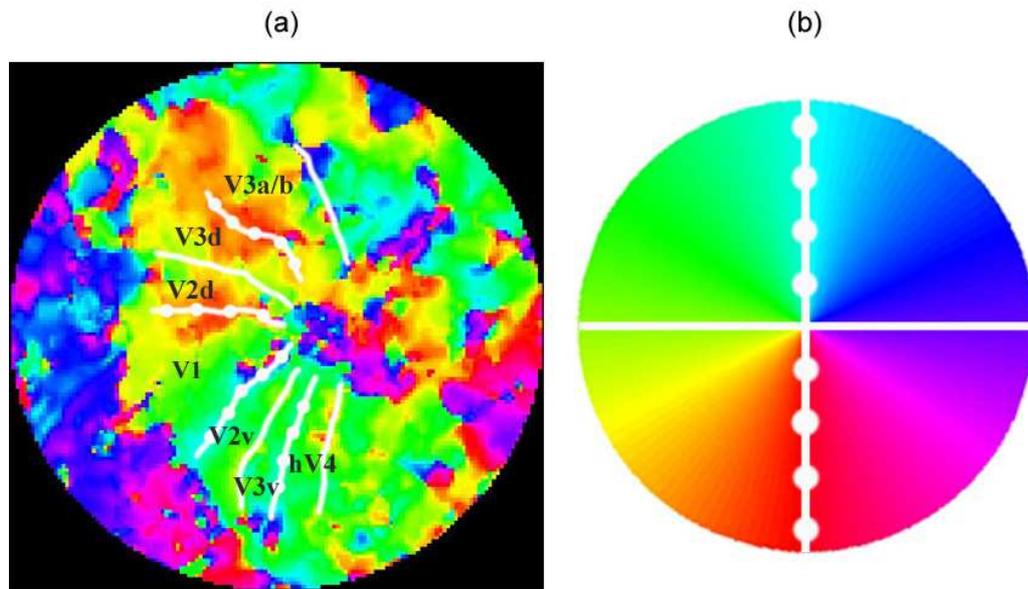


Figure 14. Determining borders of visual areas: (a) unthresholded flattened polar angle phase map of right hemisphere, (b) color legend relating phase to location of stimulus in visual field

horizontal meridian of the contralateral (opposite to hemisphere being studied) visual field. These meridian lines constitute the borders of the early visual areas (V1, V2, V3, V4). The clarity and confidence with which we can mark these borders depends upon several factors including most significantly: the fidelity of the gray-white matter segmentation that was the source of the flattened mesh, the accuracy of the alignment parameters coregistering functional and anatomical datasets, and our ability to distinguish signal from noise. It is the final factor of signal detection on which we have focused in the research presented here. Assuming mesh creation and coregistration to functional data have been performed optimally, our ability to draw borders with confidence depends on the contiguous spatial extent of significantly active regions.

Given a dataset that we believe has been optimized for the first two conditions and exhibits the retinotopic patterns described above, as a method for qualitative corroboration in a real fMRI study of our previous findings on performance and validity, we visually compare the spatial extent of active regions determined by empirical thresholding of Co , FPQ , and MSC . Using a null dataset, empirical $p = 0.05$ thresholds were determined by computing separately the Co , FPQ^* , FPQ , and MSC statistics for each of 80,000 voxels in the dataset, ordering values for each statistic in sets of 1000, and taking the mean of the values demarcating the top 5%-cutoff across the 80 sets. We chose empirical thresholding to minimize to the best of our ability bias in inference.

6.2. Results

Figure 15 shows, on flattened representations of a subject's right hemisphere, the results of thresholding the three signal detection statistics Co , FPQ (and its variant FPQ^*), and MSC according to their empirically determined $p = 0.05$ thresholds. Since visual areas are defined by identification of contiguous phase representations of the contralateral visual field, using the unthresholded map, we drew a black line around the extent of the region that would be used to label visual areas if statistical significance were not taken into consideration (Fig. 15(a)). The superimposed black border was projected onto the thresholded maps (Figures 15(b-e)) to serve as a reference for the spatial extent of potentially usable retinotopically organized regions.

Although the exact distribution of true positive activation is unknown, we can

see that most of the extent of the retinotopically organized region remains for all the statistics. The differences that do exist, suggests that some or all of them are biased from an actual significance level of $\alpha = 0.05$. Given findings using ROC curves, the validation tests, and the fact that *MSC* has produced the most conservative yet not overly-conservative activation map in Fig. 15, we conclude that *MSC* is the safest choice for thresholding activation maps in periodic fMRI studies. While borders can always be drawn without the use of a thresholded map, accurate reporting of the confidence with which borders have been identified, will depend on the use of unbiased thresholded statistics.

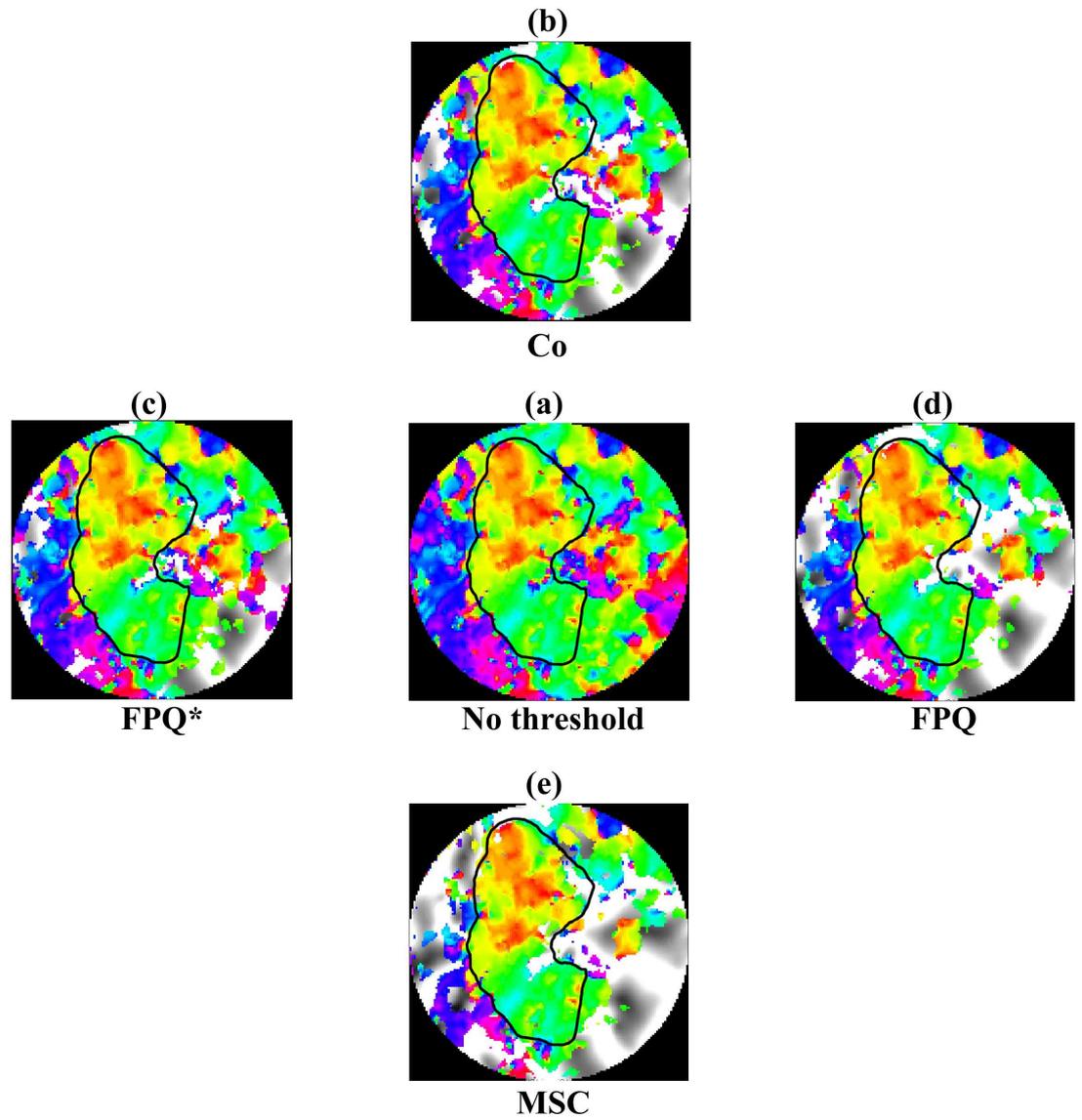


Figure 15. Flattened polar angle phase maps of right hemisphere computed at fundamental frequency and thresholded for $p = 0.05$ using empirically determined distributions of each statistic

7. DISCUSSION

While fMRI studies of periodic designs are often used for the distinct advantage of the frequency domain signal detection statistics available to them, the various statistics themselves have specific features that are responsible for variability in performance and inferential validity. In an effort to improve the sensitivity of retinotopic mapping in the human visual cortex, we compared both the performance and validity of the three statistics: Co , FPQ , and MSC using simulated and real data.

Widely used in the mrVista toolset for retinotopic mapping, Co proved to have greatest performance in ROC curve analysis using real and realistically simulated datasets (Figures 3, 6, and 7). The coincidence of its curves with the independence-assuming FPQ^* statistic and separation from FPQ , which took autocorrelation of residual terms into account, flagged a potential bias in performance estimation due to an independence assumption. MSC , computed from only information at the paradigm frequency, showed minimal variability in response to phase-coherent noise at non-paradigm frequencies (Fig. 4) of simulated data and exhibited coincidence with FPQ in real data (Fig. 7), confirming the advantage of single frequency dependence and proposing a possible robustness to assumptions of independent error terms respectively.

We further explored the influence of assumptions of independent error terms and theoretical distributions on the validity of inferences drawn from all three statistics. The assumption of independent error terms in signal detection proved to be

responsible for bias in Co and FPQ^* for low frequency paradigms studies such as those used in retinotopy (Fig. 9). Results of both the simulated and real data suggested that residual autocorrelation should be accounted for by the AR(1) modeling implemented for FPQ or with the use of MSC that showed to be robust to the independence assumption. The assumption of theoretical distributions for thresholding proved to exhibit either positive or negative bias in inference for each of the related distributions (Fig. 10) suggesting the use of empirical distributions when possible, or otherwise the more conservative theoretical distributions: *Chi*-squared for FPQ and *F*-stat for MSC .

Finally, we considered results of thresholding real human retinotopic datasets based on empirical distributions of each statistic. Because it is impossible to know the exact distribution of truly active voxels in a real dataset, defining a region that exhibited retinotopic behavior, we visually compared the extent of remaining activation in that region for $p = 0.05$ thresholded values of Co , FPQ^* , FPQ , and MSC . While Co and both FPQ s demonstrated larger overall coverage, MSC displayed the least activation outside of the region of interest and most of the activation inside the region. The confidence with which retinotopic borders are drawn depends on accurate statistical inference. Thus, based on simulations and preliminary results of real data, we recommend that when inferential statistics are in order the more conservative yet still useful statistic of MSC should be used.

Although retinotopy was the application of focus in this thesis, these results should extend to other low-frequency periodic fMRI paradigm studies. In particular, for

those applications (unlike retinotopy) in which large contiguous regions of activation are not required, detection of sinusoidal activation may more confidently be reported with *MSC* without loss in utility.

While we have met the objectives of our research as stated summarized above, the evaluation of the relative performance and validity of *Co*, *FPQ*, and *MSC* did not take into consideration the influence of EPI distortion and the multiple comparisons problem. Because of the use of gradient echo pulse sequences, magnetic susceptibility differences will cause distortions in the EPI images at tissue-bone or tissue-air interfaces. Although we believe the effect is minimal for our real retinotopy datasets, it can be a potentially significant problem in the some brain regions and with some pulse sequences and should, if possible, be addressed by applying distortion correction algorithms during preprocessing. The results of our analysis can also be further refined by considering the influence of the multiple comparisons problem. When large numbers of statistical tests are performed (such as the many thousands of voxels tested in an fMRI volume), there is potential of increased false-positive rates. Whether or not this is case for the three statistics studied here, it would be useful to evaluate the effect of various multiple comparisons correction methods on actual significance levels.

APPENDIX A. HEMODYNAMIC MODELING OF SUPERIMPOSED SIGNAL

A.1. Balloon model

We used a somewhat sophisticated model of the hemodynamic response to simulate the BOLD response time series to a stimulus. A complete model of the hemodynamic response as measured with BOLD-fMRI would involve the physiology of arteries, capillaries, and veins. BOLD modeling is simplified, however, by the observation that BOLD-fMRI is weighted heavily toward veins.³⁹ The model proposed by Buxton and used here is known as the *balloon model*.⁴⁰ It relates venous blood volume V , inward blood flow F_{in} , and percent deoxyhemoglobin Q through their normalized (to resting-state values V_0 , F_0 , and Q_0) dimensionless variables v , f_{in} and q .

The following differential equations relating q and v to blood flow result from the conservation of total blood and total deoxyhemoglobin in the venous (balloon) compartment:

$$\tau_0 \dot{v} = f_{\text{in}} - f_{\text{out}} \quad (14)$$

$$\tau_0 \dot{q} = f_{\text{in}} \frac{E}{E_0} - qv^{1/\alpha-1} \quad (15)$$

The time constant, $\tau_0 \equiv V_0/F_0$, is interpreted as the mean transit time across the balloon compartment, and E is the fraction of oxygen extracted from the blood flowing into the capillary bed.

Defined in terms of the relationship between inward flow f_{in} and resting state oxygen extraction fraction E_0 , E is the metabolic basis of the model and is approximated

as:

$$E = 1 - (1 - E_0)^{1/f_{\text{in}}} \quad (16)$$

E is a decreasing function of f_{in} reflecting the finding that the cerebral blood flow response to neural activity exceeds the metabolic demand for oxygen. By modeling E in this way, however, it does not reflect a direct neural coupling that has been used by some to explain a sometimes observed initial dip in the BOLD signal.⁴¹ Variability in E is considered here to be due entirely to activity-induced changes in blood flow.

Determined physically by the relative pressures between the venous compartment and the down-stream vein, a proper definition of outward flow f_{out} introduces variables that are not directly accessible. The balloon model replaces this detail with a phenomenological relationship between f_{out} and v . We used the simplest model that fits the data at steady-state⁴²:

$$f_{\text{out}} = v^{1/\alpha} \quad (17)$$

A.2. Neurovascular coupling

Since inward cerebral blood flow is understood to be caused (at least predominantly) by neural activity, f_{in} is defined with a model of neurovascular coupling. While the relationship between inward flow to the physiological variables v and q has been investigated extensively in recent literature and is considered a quite reasonably represented by the low-order balloon model, research into neurovascular coupling, has produced less reliable models. The model used here, which is the only one available

in the literature, is the second order linear model proposed by Friston et al.³³:

$$\ddot{f}_{\text{in}} = \epsilon u(t) - \frac{1}{\tau_s} \dot{f}_{\text{in}} - \frac{1}{\tau_f} (f_{\text{in}} - 1) \quad (18)$$

The input $u(t)$ is flow-inducing neuronal activity and as in most fMRI studies is approximated by the stimulus waveform.

A.3. MRI physics

Finally, once the outputs of the balloon model (the time series for q and v) have been computed, the BOLD response $B(t)$ is predicted using a simple linear relationship:

$$B(t) = V_0[k_1(1 - q) + k_2(1 - v)] \quad (19)$$

Equation 19 is a corrected version⁴³ of an earlier nonlinear form containing a q/v term. The parameters k_1 and k_2 are derived by considerations in MR physics, including a generalization of some quantities from 1.5T to 3T.⁴³ While errors in estimating these parameters result in a different relative weighting of q and v , they do not change their time courses.

A.4. Model parameters

The balloon model parameters were taken from the literature and are $E_0 = 0.4$, $\tau_0 = 3$, and $V_0 = 0.03$ from Buxton's 1998 paper⁴⁰ and $\alpha = 0.38$ from Buxton's 1997 paper.³² The neurovascular coupling model parameters were taken from Friston et al., where they were fit to brief stimuli with variable inter-stimulus interval.³³ It is not obvious if these latter parameters are optimal or even valid for our periodic

stimuli, but they are used here as a starting point. The exact values used are: $\epsilon = 0.5$, $\tau_s = 1.54$, and $\tau_f = 2.46$.

APPENDIX B. SIMULATING AUTOCORRELATIVE NOISE PROCESS

To simulate autocorrelative noise processes we implemented a f^{-1} scaling of a white noise power spectrum as proposed by Rangarajan and Ding.³⁴ Their steps are summarized here.

First, a discrete zero mean white Gaussian noise process $\{\xi_k\}$, $k=0,1,2,\dots,N-1$ is Fourier transformed to obtain:

$$\Gamma_k = \sum_{n=0}^{N-1} \xi_n e^{-i2\pi k n/N} \quad (20)$$

Next, Γ_k is multiplied by $f^{-1} = (k/N)^{-1/2}$ resulting in the scaled quantity Γ'_k . This scaled quantity is then inverse Fourier transformed to obtain the desired discrete time series, x_n :

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} \Gamma'_k e^{2\pi k n/N} \quad (21)$$

where $n=0,1,2,\dots,N-1$.

APPENDIX C. RELATIONSHIP OF CO, FPQ, AND MSC TO THEORETICAL DISTRIBUTIONS

Relationship of Co to $erfc$ (mrVista code, Bandettini et al.⁸):

$$p - value \sim (1 - erf(Co * \sqrt{N/2})) \quad (22)$$

where N follows as above.

Relationship of Co to t -statistic (Silver et al.³⁵):

$$t \sim Co \sqrt{\frac{N-2}{1-Co^2}} \quad (23)$$

Relationship of FPQ to χ^2 statistic (Bullmore et al.³):

$$\chi^2 \sim 2(FPQ) \quad (24)$$

Relationship of MSC to F -statistic (Simpson et al.²⁶):

$$F_{2,2M-2} \sim \frac{MSC(M-1)}{1-MSC} \quad (25)$$

where M is the total number of time series samples in the estimate as defined previously. P-values for t , χ^2 , and F values can subsequently be assigned from their respective distribution tables.

REFERENCES

1. D. Felleman and D. V. Essen, "Distributed hierarchical processing in the primate cerebral cortex," *Cerebral Cortex* **1**, pp. 1–47, 1991.
2. S. Engle, G. Glover, and B. Wandell, "Retinotopic organization in human visual cortex and the spatial precision of functional MRI," *Cerebral Cortex* **7**, pp. 181–192, 1997.
3. E. Bullmore, M. Brammer, S. Williams, S. Rabe-Hesketh, N. Janot, A. David, J. Mellers, R. Howard, and P. Sham, "Statistical methods of estimation and inference for functional MR image analysis," *Magnetic Resonance in Medicine* **35**, pp. 261–277, 1996.
4. S. Ogawa, T. Lee, A. Nayak, and P. Glynn, "Oxygenation-sensitive contrast in magnetic resonance image of rodent at high magnetic fields," *Magnetic Resonance in Medicine* **14**, pp. 68–78, 1990.
5. S. Ogawa, T. Lee, A. Kay, and D. Tank, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation," *Proc. Natl. Acad. Sci.* **87**, pp. 9869–9872, 1990.
6. S. Ogawa, D. Tank, R. Menon, J. Ellermann, S. Kim, H. Merkle, and K. Ugurbil, "Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging," *Proc. Natl. Acad. Sci.* **89**, pp. 5951–5955, 1992.
7. P. Bandettini, E. Wong, R. Hinks, R. Tikofsky, and J. Hyde, "Time course EPI in human brain function during task activation," *Magnetic Resonance in Medicine* **25**, pp. 390–397, 1992.
8. P. Bandettini, A. Jesmanowicz, E. Wong, and J. Hyde, "Processing strategies for time-course data sets in functional MRI of the human brain," *Magnetic Resonance in Medicine* **30**, pp. 161–173, 1993.
9. K. Friston, A. Holmes, J. Poline, P. Grasby, S. Williams, and R. Frackowiak, "Analysis of fMRI time series revisited," *Neuroimage* **2**, pp. 45–53, 1995.
10. G. Boynton, S. Engel, G. Glover, and D. Heeger, "Linear systems analysis of functional magnetic resonance imaging in human V1," *Journal of Neuroscience* **16**, pp. 4207–4221, 1996.
11. M. Cohen, "Parametric analysis of fMRI data using linear systems methods," *Neuroimage* **97**, pp. 93–103, 1997.

12. G. Aguirre, E. Zarahn, and M. D'Esposito, "The variability of human, BOLD hemodynamic response," *Neuroimage* **8**, pp. 360–369, 1998.
13. R. Buckner, P. Bandettini, K. O'Craven, R. Savoy, S. Petersen, M. Raichle, and B. Rosen, "Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging," *Neurobiology* **93**, pp. 14878–14883, 1996.
14. M. Burock, R. Buckner, M. Woldorff, B. Rosen, and A. Dale, "Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI," *Neuroreport* **9**, pp. 3735–3739, 1998.
15. B. Rosen, R. Buckner, and A. Dale, "Event-related functional MRI: past, present, future," *Proc. Natl. Acad. Sci.* **95**, pp. 773–780, 1998.
16. A. Dale, "Optimal experimental design for event-related fMRI," *Human Brain Mapping* **8**, pp. 109–114, 1999.
17. A. Dale and R. Buckner, "Selective averaging of rapidly presented individual trial using fMRI," *Human Brain Mapping* **5**, pp. 329–340, 1997.
18. T. Oakes, T. Johnstone, K. O. Walsh, L. Greischar, A. Alexander, A. Fox, and R. Davidson, "Comparison of fMRI motion correction software tools," *Neuroimage preprint*, 2005.
19. K. Worsley and K. Friston, "Analysis of fMRI time-series revisited - Again," *NeuroImage* **2**, pp. 173–182, 1995.
20. E. Zarahn, G. Aguirre, and M. D'Esposito, "Empirical analyses of BOLD fMRI statistics: I. Spatially unsmoothed data collected under null-hypothesis conditions," *NeuroImage* **5**, pp. 179–197, 1996.
21. P. Purdon and R. Weisskoff, "Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI," *Human Brain Mapping* **6**, pp. 239–249, 1998.
22. J. Locascio, P. Jennings, C. Moore, and S. Corkin, "Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance brain imaging," *Human Brain Mapping* **5**, pp. 168–193, 1997.
23. M. Woolrich, B. Ripley, M. Brady, and S. Smith, "Temporal autocorrelation in univariate linear modeling of fMRI data," *NeuroImage* **14**, pp. 1370–1386, 2001.

24. J. Marchini and S. Smith, "On bias in the estimation of autocorrelations for fMRI voxel time-series analysis," *NeuroImage* **18**, pp. 83–90, 2003.
25. E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T. Carpenter, and M. Brammer, "Colored noise and computational inference in neurophysiological (fMRI) time series analysis: resampling methods in time and wavelet domains," *Human Brain Mapping* **12**, pp. 61–78, 2001.
26. D. Simpson, C. Tierra-Criollo, R. Leite, E. Zayen, and A. Infantosi, "Objective response detection in an electroencephalogram during somatosensory stimulation," *Annals of Biomedical Engineering* **28**, pp. 691–698, 2000.
27. R. Dobie and M. Wilson, "Analysis of auditory of evoked potentials by magnitude squared coherence," *Ear and Hearing* **10**, pp. 2–13, 1989.
28. R. Constable, P. Skudlarski, and J. Gore, "An ROC approach for evaluating functional brain MR imaging and postprocessing protocols," *Magnetic Resonance in Medicine* **34**, pp. 57–64, 1995.
29. J. Sorenson and X. Wang, "ROC methods for evaluation of fMRI techniques," *Magnetic Resonance in Medicine* **36**, pp. 737–744, 1996.
30. P. Skudlarski, R. Constable, and J. Gore, "ROC analysis of statistical methods used in functional MRI: individual subjects," *NeuroImage* **9**, pp. 311–329, 1999.
31. R. Nandy and D. Cordes, "Novel ROC-type method for testing the efficiency of multivariate statistical methods in fMRI," *Magnetic Resonance in Medicine* **49**, pp. 1152–1162, 2003.
32. R. Buxton and L. Frank, "A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation," *Journal of Cerebral Blood Flow and Metabolism* **17**, pp. 64–72, 1997.
33. K. Friston, A. Mechelli, R. Turner, and C. Price, "Nonlinear responses in fMRI: The balloon model, Volterra kernels, and other hemodynamics," *Neuroimage* **12**, pp. 466–477, 2000.
34. G. Rangarajan and M. Ding, "Integrated approach to the assessment of long range correlation in time series data," *Physical Review E* **61**, pp. 4991–5001, 2000.
35. M. Silver, D. Ress, and D. Heeger, "Topographic maps of visual spatial attention in human parietal cortex," *Journal of Neurophysiology* **94**, pp. 1358–1371, 2005.

36. S. Engle, D. Rumelhart, B. Wandell, A. Lee, G. Glover, E. Chichilnisky, and M. Shadlen, “fMRI of human visual cortex,” *Nature* **369**, p. 525, 1994.
37. M. Sereno, A. Dale, J. Reppas, K. Kwong, J. Belliveau, T. Brady, B. Rosen, and R. Tootell, “Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging,” *Science* **268**, pp. 889–893, 1995.
38. K. Kwong, J. Belliveau, D. Chesler, I. Goldberg, R. Weisskoff, B. Poncelet, D. Kennedy, B. Hoppel, M. Cohen, and R. Turner, “Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation,” *Proc. Natl. Acad. Sci.* **89(12)**, pp. 5675–5679, 1992.
39. R. Buxton, *Introduction to Functional Magnetic Resonance Imaging*, Cambridge University Press, Cambridge, United Kingdom, 2002.
40. R. Buxton, E. Wong, and L. Frank, “Dynamics of blood flow and oxygenation changes during brain activation: the balloon model,” *Magnetic Resonance in Medicine* **39**, pp. 855–864, 1998.
41. T. Ernst and J. Hennig, “Observation of a fast response in functional MR,” *Magnetic Resonance in Medicine* **32**, pp. 146–149, 1994.
42. R. Grubb, M. Raichle, J. Eichling, and M. Ter-Pogossian, “The effects of changes in PaCO₂ on cerebral blood volume, blood flow, and vascular mean transit time,” *Stroke* **5**, pp. 630–639, 1974.
43. T. Obata, T. Liu, K. Miller, W.-M. Luh, E. Wong, L. Frank, and R. Buxton, “Discrepancies between BOLD and flow dynamics in primary and supplementary motor areas: application of the balloon model to the interpretation of BOLD transients,” *Neuroimage* **21**, pp. 144–153, 2004.