

2002-09

# ROI-Based Analysis of Functional Imaging Data

---

<https://hdl.handle.net/2144/2298>

*"Downloaded from OpenBU. Boston University's institutional repository."*

## **ROI-Based analysis of functional imaging data**

Alfonso Nieto-Castanon, Satrajit Ghosh, Jason Tourville and Frank Guenther

**September, 2002**

**Technical Report CAS/CNS-2002-010**

Permission to copy without fee all or part of this material is granted provided that: 1. The copies are not made or distributed for direct commercial advantage; 2. the report title, author, document number, and release date appear, and notice is given that copying is by permission of the BOSTON UNIVERSITY CENTER FOR ADAPTIVE SYSTEMS AND DEPARTMENT OF COGNITIVE AND NEURAL SYSTEMS. To copy otherwise, or to republish, requires a fee and / or special permission.

Copyright © 2002

Boston University Center for Adaptive Systems and  
Department of Cognitive and Neural Systems  
677 Beacon Street  
Boston, MA 02215

# ROI-Based Analysis Of Functional Imaging Data

Alfonso Nieto-Castanon<sup>\*</sup>, Satrajit S. Ghosh<sup>\*</sup>, Jason A. Tourville<sup>\*\*†</sup>, Frank H. Guenther<sup>\*\*‡</sup>

<sup>\*</sup> Department of Cognitive and Neural Systems, Boston University

<sup>†</sup> Center for Morphometric Analysis, Massachusetts General Hospital

<sup>‡</sup> Research Laboratory of Electronics, Massachusetts Institute of Technology

Keywords: fMRI; MRI; methods; region of interest; anatomical variability; multivariate analysis

## Abstract

In this technical report, we present fMRI analysis techniques that test functional hypotheses at the region of interest (ROI) level. An SPM-compatible Matlab toolbox has been developed which allows the creation of subject-specific ROI masks based on anatomical markers and the testing of functional hypotheses on the regional response using multivariate time-series analysis techniques. The combined application of subject-specific ROI definition and region-level functional analysis is shown to appropriately compensate for inter-subject anatomical variability, offering finer localization and increased sensitivity to task-related effects than standard techniques based on whole brain normalization and voxel or cluster-level functional analysis, while providing a more direct link between discrete brain region hypotheses and the statistical analyses used to test them.

## 1 INTRODUCTION

Most hypotheses that are addressed using functional magnetic resonance imaging (fMRI) are stated in terms of the specific functionality of brain regions of interest (ROIs). These regions are frequently defined based on their cytoarchitectonic structure (e.g. Brodmann areas) or anatomical landmarks such as sulci (Rademacher et al. 1993; Caviness et al. 1996). It is widely acknowledged (though rarely measured) that there exists a considerable degree of inter-subject variability in the shape and location of these regions. We begin this paper by presenting evidence that standard normalization techniques only partially accommodate inter-subject variability, and that after a full-brain normalization procedure there exists a considerable degree of residual variability in the shape and location of regions defined based on anatomical markers.

Since most fMRI experiments are built on multiple-subject data, standard functional analysis techniques based on voxel-level statistics attempt to compensate for this variability by spatially smoothing the functional series after normalization. Smoothing at the same time enforces the validity of standard statistical analysis based on Gaussian field theory (Friston et al. 1996). A troubling aspect of this solution is the loss of clear regional boundaries resulting from the smoothing of the BOLD response across neighboring but possibly functionally dissimilar regions. Furthermore, the sensitivity of the resulting statistical tests is expected to decrease with the extent of anatomical variability across subjects.

In the present work we take a different strategy and present a methodology for the analysis of functional data that focuses on the activation of specific brain regions of interest. The proposed methodology is based on the definition of subject-specific ROIs and the testing of functional hypotheses directly at the level of the (multivariate) whole region activation. In this way, we avoid the need of full-brain inter-subject coregistration and, most importantly, the need of spatially smoothing the functional series. By providing confirmatory analyses at the level of ROIs, the proposed methodology serves as a more direct link between the initial research hypotheses stated in terms of the functionality of discrete brain regions and the functional analyses used to test these hypotheses. This confirmatory approach to regional functional analysis is expected to ultimately increase the replicability of fMRI experiments and facilitate a knowledge buildup from functional results.

The outline of the paper is as follows. Section 2 describes a tool for the definition of ROIs based on anatomical markers, and illustrates the extent of inter-subject anatomical variability in temporal lobe cortical regions on a set of nine subjects. Section 3 summarizes the proposed methodology for the functional analysis of regional imaging data. Simulations comparing the expected sensitivity of the proposed methodology to one based on inter-subject full-brain normalization and voxel- or cluster-level analyses are presented in Section 4. Last, Section 5 presents Monte Carlo simulations validating the proposed statistical functional analyses on a range of possible fMRI noise conditions.

## 2 INTER-SUBJECT ANATOMICAL VARIABILITY

### Definition of subject-specific regions of interest

A Matlab-based interactive toolbox was developed for the parcellation of ROIs from structural MRI scans. This toolbox currently allows for semi-automated ROI identification based on anatomical markers and is available for free download at <http://cns.bu.edu/~speech/> along with the functional analysis tools described in the next section. While the exact relationship

between anatomical markers and brain functionality is still open to discussion, anatomical markers provide a reliable and replicable ROI definition and a more flexible basis for coregistration than standard global brain anatomy normalization techniques. The tools for parcellation, along with the functional analysis tools described below, were designed to interface with the SPM functional imaging analysis package (<http://www.fil.ion.ucl.ac.uk/spm>).

A systematic labeling comprises the following steps:

**Cortical Segmentation.** T1-weighted images are preprocessed to extract white matter-gray matter and gray matter-CSF boundaries. The boundaries, which are represented as contours, form a cortical ribbon segmentation. The preprocessed images are then manually edited in the coronal plane to correct segmentation errors. The user can interactively adjust the isocontour levels along the coronal dimension to correct any automatic segmentation errors.

**Sulci and Node Identification.** ROI-limiting sulci are identified and traced on the three canonical image planes. Nodal points are demarked at relevant sulcus intersections and cortical/subcortical landmarks. In general, sulci form ROI medio-lateral boundaries while nodal points mark their rostral-caudal extents.

**Parcellation.** Based on limiting fissures and nodal points, the segmented cortical ribbon is parcellated into smaller units. The parcellation procedure is based on methods developed at the Center for Morphometric Analysis at the Massachusetts General Hospital (see Rademacher et al. 1993; Caviness et al. 1996).

**Labeling.** Each parcellated unit is assigned a label based upon its position relative to the nodal points. The labeled cortical regions constitute the ROI masks used for functional analysis.

#### Demonstration of anatomical variability

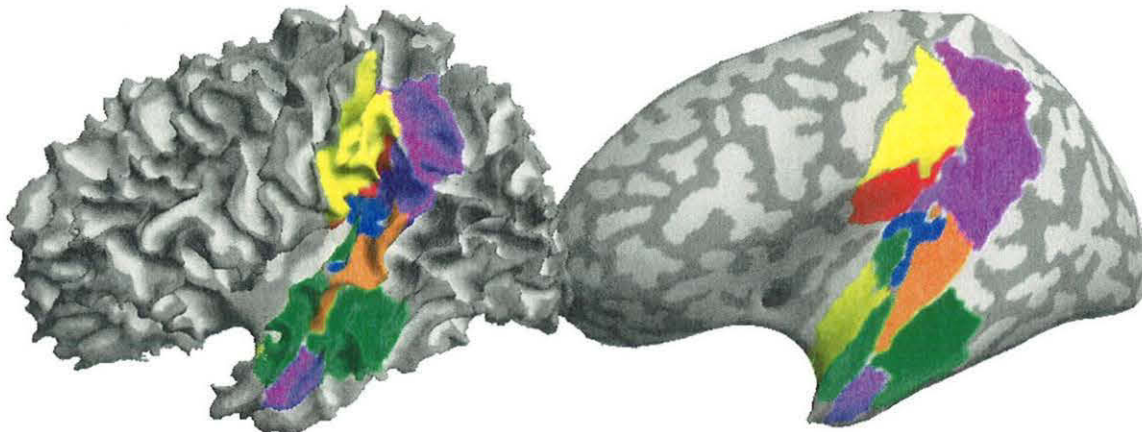
To demonstrate the extent of local anatomical variability, ten ROIs in each hemisphere corresponding to perisylvian cortical areas in the temporal and parietal lobes were parcellated and labeled by hand for nine subjects using their T1-weighted MRI scans. SPM normalization (full normalization including affine and 7x8x7 nonlinear basis functions) was then applied to each of these images, and the extent of variability in the locations of the original regions in the resulting normalized space was analyzed.

The ROIs are shown in Figure 1 in two separate formats. The left panel shows the ROIs overlaid on a brain surface that represents the gray-white boundary. The right panel surface results from inflating the surface on the left. The inflation improves the clarity of observing the ROIs on the superior temporal plane. Performing the overlay also provides a method of testing the accuracy of the parcellation and comparison of inter-parcellator reliability.

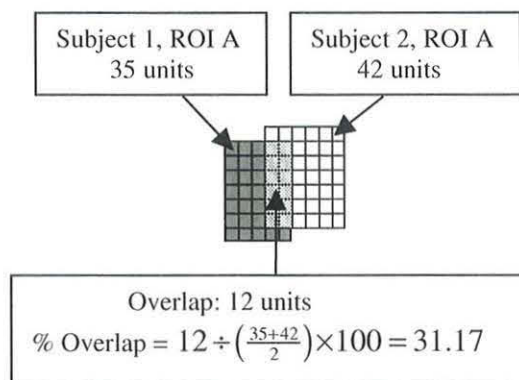
The original T1 scans (voxel size: 1.33x1x1 mm) were rotated using a transformation that aligned the AC-PC line perpendicular to the coronal plane. Gray matter around the temporal lobe was parcellated in each of these volumes using a definition of regions as described in Caviness et al. (1996). These MR volumes and the corresponding ROI definitions were then normalized using the SPM package. The volumes were resampled to have isotropic voxels of size 2x2x2 mm. The normalized structural images and ROI definitions were then manually checked for coregistration.

For each subject, the volume of an ROI was determined by counting the total number of voxels belonging to that ROI. Overlap percentages were calculated for all possible combinations of subjects in the test set. Thus, there are  $\binom{9}{2}$  possible combinations of two-subject pairs,

$\binom{9}{3}$  combinations of 3-subject subsets and so on. For each of the 36 two-subject pairs, the overlap for each ROI was defined as the number of voxels in the intersection of the ROI between the subjects divided by the mean number of voxels in the ROI for each subject as demonstrated in the schematic in Figure 2.



**FIGURE 1. ROI Overlay.** The left panel shows the ROIs overlaid on the gray-white surface. The right panel shows the ROIs overlaid on the inflated surface. HG = Heschl's gyrus; PT = planum temporale; PP = planum polare; T1a,T1p = anterior/posterior superior temporal gyrus; T2a,T2p = anterior/posterior middle temporal gyrus; PO = parietal operculum; SGa,SGp = anterior/posterior supramarginal gyrus. Because of its location deep in the Sylvian fissure, PP only shows up in the inflated view on the right.



**FIGURE 2. Overlap computation schema.** The overlap is calculated for the same ROI of two subjects. The overlap is calculated as the intersection of the volumes divided by the mean of the volumes. 2D meshes represent the volumes in this schematic.

This calculation is repeated for all possible combinations of subjects for each ROI. The mean overlap percentage is determined for each ROI across all possible two-subject combinations. The same is done for all 3-subject combinations, 4-subject combinations and so on. The results are shown in Figure 3. The ROIs are sorted from top to bottom in ascending order of the mean size of each ROI across all 9 subjects. The amount of overlap reduces drastically as more subjects are pooled together. For example, RightT2p (right hemisphere, posterior middle temporal gyrus) shows a mean overlap of 37.47% across 2-subject combinations. The mean overlap drops to 18.78% for 3-subject combinations and to 1.43% for all nine subjects. This trend is similar across all the ROIs considered.

### 3 FUNCTIONAL ANALYSIS

The relatively large extent of anatomical variability demonstrated above raises concerns about the statistical power and replicability of functional analyses based on whole-brain normalization. A more robust approach for testing functional hypotheses would be to define subject-specific ROIs expressing the focus of the researcher's interest while accommodating inter-subject variability in the expected loci of activation, and then test the specific functional hypotheses on the activation of voxels lying within these ROIs. This section describes statistical tests tailored to region-level functional analysis.

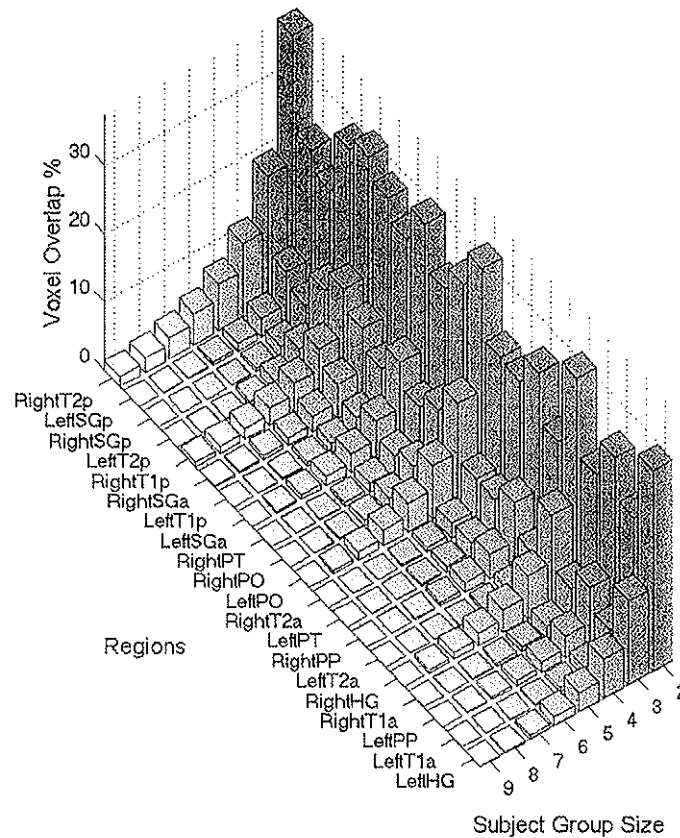


FIGURE 3. Mean overlap for each ROI across different subject group sizes. The ROIs are subdivided as belonging to left and right hemispheres. For example, Right HG (right hemisphere Heschl's gyrus) shows a mean overlap of 31.43% across 2-subject combinations. The mean overlap drops to 13.26% for 3-subject combinations and to no overlap for all nine subjects.

The functional analysis techniques described here are founded upon previous proposals for multivariate analysis of fMRI data (Friston et al. 1995; Worsley et al. 1997). These proposals focus on the analysis of whole-brain activation. In the present work we are interested in the analysis of specific brain regions. This has important consequences in the type of analysis technique chosen since some of the assumptions valid at the level of the whole brain are not appropriate for smaller regions, and vice-versa. Friston et al. (1995) propose a multivariate treatment of whole-brain activation that does not make any explicit assumptions about the spatial

correlation in fMRI data, but it uses a heuristic argument to correct for temporal correlations. This gives a biased test, increasing the false-positive rate for data acquired with relatively small repetition times (Worsley et al. 1997). Worsley et al. (1997) presents a multivariate technique which takes into account the temporal correlation in fMRI data by estimating the true temporal degrees of freedom of the functional series, and the spatial correlation by assuming a spatially stationary Gaussian random field. The treatment of temporal correlations in this method relies on a common assumption of a spatially stationary frequency spectrum (the absolute amount of noise can vary between voxels but the shape of their frequency spectrum is stationary). This assumption is most applicable in the case of regional activations given their reduced size versus whole-brain activations. Despite the merits of this approach, the Gaussian random field assumption for modeling the spatial correlation of the data is problematic when applying these techniques to the analysis of discrete brain regions because we are forced to spatially smooth the functional data, leading to the problem of integrating the BOLD response across neighboring regions. If we were only to smooth the data within a given region, we would potentially have problems with the model validity at the regional boundaries. This is an especially challenging problem for small ROIs.

The analysis techniques proposed in this paper do not make any explicit assumptions about the spatial correlation of the data while correctly compensating for temporal correlations. Unlike Worsley et al. (1997), the treatment of temporal correlations is based on a whitening technique –similar to Purdon and Weisskoff (1998) and Worsley et al. (2002)- that permits optimal estimation of the task-related effects, and it also permits the use of multivariate statistics based on a likelihood ratio test (LRT). As proposed by Friston (Friston et al. 2000), the data is band-pass filtered prior to hypothesis testing with important consequences for the test robustness against model misspecification. The details of the proposed statistical analysis techniques are described below.

Prior to all analyses, the functional data is preprocessed at the whole-brain level (realignment of temporal series and coregistration with structural data). Using the pre-specified ROI masks, activation profiles for all voxels within each ROI are then extracted. Unless otherwise stated, all steps described in the following sections are applied to each ROI independently.

#### **Initial preprocessing steps**

Standard parametric statistical analyses test the presence of an effect of interest (such as a task-related activation response) in a temporal series by comparing the “size” of the effect (explained variability) to the “size” of the noise (unexplained variability). Valid tests are constructed by using an appropriate characterization of the noise statistical properties. The initial preprocessing steps aim at providing a characterization of the temporal and spatial correlation structure of the noise, by transforming the original functional data such that the noise follows a simple known distribution (in this case, multivariate white noise with rank-limited spatial covariance). First, the noise temporal correlation is fitted using a broad model of the fMRI BOLD signal noise and the data is “whitened” (the time series frequency components are divided by the estimated noise spectrum). Second, a data reduction step is applied to the functional data within a region so that the dimensionality of the spatial covariance is reduced to a predefined value.

*Modeling noise temporal correlation.* The regional time series are initially fitted using the standard general linear model with the given set of predictors. Residuals are used as an initial

estimation of noise. fMRI noise is modeled as a mixture of a low-frequency (Gaussian-shaped autocorrelation with unknown autocorrelation width) noise term and a (wideband) white noise term. The first term models the possible low-frequency noise components such as long term physiological shifts and movement-related noise not otherwise accounted for, while the second term models the subject and scanner thermally generated noise. The modeled noise average frequency spectrum takes the form:

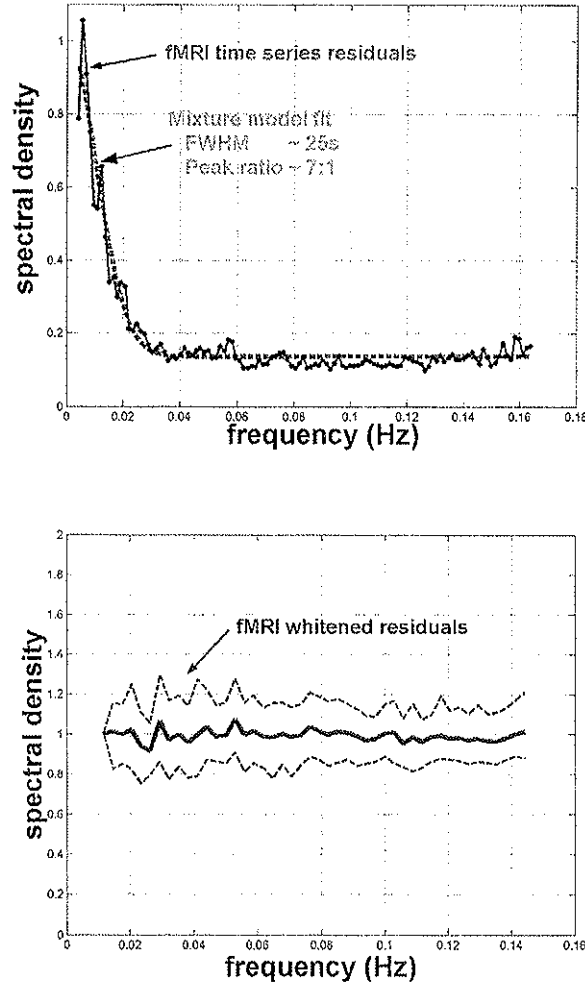
$$N(w) = a_1 e^{-\frac{w^2}{2\sigma^2}} + a_2$$

where  $N(w)$  represents the noise spectral energy at frequency  $w$ , and  $a_1$ ,  $a_2$ , and  $\sigma$  are free model parameters (the ratio  $a_1/a_2$  measures the relative contribution of each noise term and is referred to in this paper as *peak ratio*). An Expectation-Maximization (EM) algorithm is used to estimate the three free model parameters for each ROI. The algorithm has been modified to estimate and correct for the part of the modeled noise that correlates with the predictors and would therefore be missing from the obtained residuals. Finally each voxel is whitened using the fitted noise spectrum and band-pass filtered using a pre-specified frequency of interest rectangular window.

Figure 4 (top) shows an example of the observed BOLD signal noise frequency spectrum and the corresponding model fit. The noise spectrum was estimated from fMRI residuals (EPI at 1.5T, 170 scans, TR=2s, motion-corrected and detrended, in a passive listening condition block design) from a region of 396 voxels that was not found to respond to the experimental paradigm using standard SPM analyses. Other noise models, such as the  $1/f$  model (Zarahn et al. 1997) or AR(n) models (Bullmore et al. 1996) have been shown to produce valid fits to the estimated spectral energy of fMRI residuals. Like the  $1/f$  model (of the form  $a_1 + a_2/(w + a_3)$ ), the proposed noise model has the advantage of a reduced set of free parameters. The use of this particular model herein is based on the observation (Friston et al. 2000) that comparatively good fits of  $1/f$  models are partly due to linear and movement related trends in the BOLD signal. Once these trends are removed, we find the remaining noise spectral shape to be well approximated by the proposed mixture model. To validate the use of this particular noise model we computed the noise spectrum from 20 perisylvian cortical regions in 6 subjects (same conditions as the example above). The average spectrum of the residuals was fitted individually for each region and whitened using the estimated mixture model spectrum. Results are shown in Figure 4 (bottom). The whitened series do not show any visible frequency variation in the spectral density, confirming that the resulting whitened residuals would be well approximated as white noise.

*Data Reduction.* To allow a valid estimation of the noise multivariate spatial covariance, the dimensionality of the signal (originally the number of voxels within an ROI) must be reduced. This is accomplished by projecting each regional response onto an orthogonal spatial basis and keeping only a fixed number of the resulting components. For the definition of the spatial basis we used two alternative methods. The first method, similar to Friston et al. (1995), produces a signal-dependent spatial basis using a singular value decomposition (SVD) of the regional response, only keeping the first few (maximal variance) components. The second defines a signal-independent spatial Fourier basis and keeps only the first few (low-spatial-frequency) components for each spatial dimension. Projecting the regional responses onto an orthogonal spatial basis produces a reduced set of temporal eigenvariates which are then used for hypothesis testing. The spatial bases are chosen in an effort to obtain resulting eigenvariates that comprise most of the signal and a limited amount of the noise covariance. A signal-dependent basis is

expected to best accomplish this noise reduction if a relatively large effect size can be assumed since it makes no assumptions about the signal spatial covariance. A signal-independent basis would work best for small effect sizes –where an SVD decomposition would fail to pick the signal components– when relatively smooth spatial activation profiles can be assumed. All the results shown in the present paper are obtained using a signal-independent spatial basis for data reduction.



**FIGURE 4.** Top: Example of fMRI noise frequency spectrum (solid line), and the corresponding mixture model fit (dashed line). Physiological noise autocorrelation width is estimated to be 25s full-width-half-maximum, with a relative contribution of 7:1 between the physiological (low-frequency) and thermal (wideband) noise components. The estimated noise spectrum is used to whiten the functional series. Bottom: Validation of mixture noise model. Solid line shows the average frequency spectrum of the residuals from fMRI data (EPI at 1.5T, TR=2s) after whitening using the estimated mixture model. Dashed lines show the 5% and 95% percentiles of the whitened frequency spectrum. Results are pooled for 20 regions in 6 subjects. The noise model is fitted individually for each region/subject.

### General linear model and hypothesis testing

Once the data has been preprocessed, the resulting activation profiles are modeled using the general linear model. As in standard SPM analyses, the researcher specifies the form of the task-related effects on the temporal activation profiles, and a test on a specific contrast defining the presence or the comparative strength of these effects is provided. Contrasting with the SPM voxel-level analyses, these tests are performed on the regional activations, and the results are provided for each defined ROI. Several ways of characterizing and testing the spatial profile of the effects within each region are also provided. The details of the statistical analysis follow.

The general linear model is used to estimate the matrix of regressors fitting the resulting set of temporal eigenvariates with a given set of predictors. The model takes the usual linear form:

$$Y = X \cdot B + Z$$

where  $Y$  represents the activations within an ROI after the initial pre-processing steps (whitening and data reduction) with one column for each eigenvariate and one row for each scan.  $X$  is the design matrix (filtered with the inverse of the estimated noise spectrum) with one column for each effect and one row for each scan.  $Z$  represents zero mean white noise with unknown spatial covariance. The matrix of regressors  $B$  is estimated using Optimal Least Squares. Hypothesis testing uses a likelihood ratio test (LRT) (Mardia et al. 1979). For a simple contrast vector  $c_i$  with one element for each effect, the regional statistic used will be:

$$\lambda_F = \frac{c_i^T B (E^T E)^{-1} B^T c_i}{c_i^T (X^T X)^{-1} c_i}$$

where  $E$  represents the matrix of residuals ( $E \equiv Y - XB$ ). Under the null hypothesis ( $c_i^T \cdot B = 0$ ), this statistic follows a known  $F$  distribution with  $n$  and  $v$  degrees of freedom:

$$\lambda_F \sim \frac{n}{v} F_{n,v}$$

$$v = r - \text{rank}(X) - n + 1$$

where  $n$  is the number of eigenvariates retained and  $r$  is the number of frequency components in the frequency of interest window. This test provides a natural multivariate extension to the standard univariate chi-square test for voxel-wise contrasts (to which it reduces when one-voxel ROIs are considered).

There are two ways to characterize the spatial profile of the effects. First, the effects can be characterized by projecting the eigenvariate contrasts back onto a temporal and a spatial profile (for each contrast tested) using the transpose of the previously defined orthogonal spatial basis. The resulting spatial profiles are then averaged across subjects (simply aligning the region centers for each subject) to provide a visual display of the effect within an ROI. Second, hypothesis testing can also be performed on the spatial characterization of the effects. A spatial contrast vector, with one element for each voxel within an ROI, can be defined as an arbitrary function of each voxel's spatial position, and a T-test is provided that tests the presence of such a spatially distributed effect. For example, a spatial contrast vector containing ones in all its elements tests the average activation within a region, while a spatial contrast vector containing a

linear function on the  $y$  position of each voxel tests a differential activation along the anterior-posterior axis. The implemented spatial T-test takes the form:

$$\lambda_T = \frac{c_i^T B c_x}{\sqrt{c_x^T E^T E c_x c_i^T (X^T X)^{-1} c_i}}$$

where  $c_x$  represents the spatial contrast vector. Under the null hypothesis ( $c_i^T B c_x = 0$ ), this statistic follows a known  $T$  distribution with  $v$  degrees of freedom:

$$\lambda_T \sim \frac{1}{\sqrt{v}} T_v$$

$$v = r - \text{rank}(X).$$

This test provides a natural extension to the standard univariate T-test for voxel-wise contrasts (to which it reduces when one-voxel ROIs are considered). The data reduction preprocessing step is skipped when performing this test, as it is no longer necessary to estimate the noise multivariate spatial covariance.

While not described in this paper, the analysis tools available at <http://cns.bu.edu/~speech/> include multiple effect tests (simultaneous test of multiple contrasts or effects) similar to the tests described above for a simple contrast. These tests lead to statistics based on the  $\Lambda$  and  $F$  distributions for the described regional and spatial analyses, respectively.

**Across-subject analyses.** Fixed-effect and random-effect analyses for obtaining across-subject statistics have been implemented. For the results presented in this paper, across-subject statistics are obtained using a fixed-effect analysis framework. Standard voxel-level analyses for multiple subjects provide pooled statistics for each voxel, identifying equivalent voxels across subjects based on their spatial position after normalization. In contrast, the proposed method obtains pooled statistics for each regional eigenvariate, identifying equivalent eigenvariates across subjects based on the region they belong to and their index (1 to  $n$ ). For example, if a signal-independent basis was used for data reduction this means that we are effectively performing a regional coregistration of each subject's activation based on the spatial frequency (along each spatial direction) of the regional regressors. If a signal-dependent basis was used, then no within-region spatial coregistration is assumed across subjects, and the effects are pooled based on their relative strength for each subject.

#### 4 SIMULATIONS

To simulate the effect of the observed anatomical variability on standard functional analyses, Monte Carlo simulations were run for both a standard procedure involving inter-subject coregistration followed by voxel- and cluster-level analyses and the proposed ROI methodology to analyze controlled simulated functional data for nine subjects. The subjects' parcellated structural images were used as reference for the location of simulated fMRI data. Simulated functional runs consisted of a single run for each subject with 128 full brain acquisitions, TR=2s, voxel size 4x4x4mm. The simulated data were created as white noise subsequently temporally and spatially smoothed using a spatial-temporal Gaussian kernel with full-width-half-maximum (FWHM) of 7s and 4mm respectively. This noise activation was distributed over all defined ROIs for each of the subjects. A sinusoidal signal (T=32s) was added with a predefined signal-to-noise ratio and distributed over a portion (25% most posterior voxels) of the actual location of

left Heschl's gyrus for each of the nine subjects labeled. Signal-to-noise ratios of 1%, 5%, and 10% residual mean square (RMS) at the voxel level were used (labeled as 'small', 'medium' and 'large' effect sizes, respectively). A time series following the sinusoidal signal activation was introduced as a predictor, and hypothesis testing was performed on the level of the estimated regressor for this predictor using both methodologies. The statistical analysis centered on a 1/64 to 1/16 Hz frequency window. 50 simulations were run for each effect size tested.

**Processing steps for methodologies based on inter-subject coregistration:** Standard functional pre-processing steps included SPM inter-subject normalization (using full normalization including affine and 7x8x7 nonlinear basis functions) and spatial smoothing (Gaussian kernel, FWHM=8mm). Results were analyzed in terms of the estimated peak of activation locations, and their corresponding corrected voxel-level and cluster-level p-values (across-subject analyses were obtained using a fixed-effect framework). The location of the obtained peak of activation was labeled based on the regional identity of the corresponding voxel for each subject's normalized structural data. Results were averaged across these different subject-based regional identities.

**Processing steps for the ROI methodology:** Two statistical tests were run for each simulation: an F-test on the presence of an effect at the region level, and a T-test on the presence of a differential effect along the anterior-posterior axis within each region. The spatial contrast vector  $c_x(i)$  for the T-test was defined for each region as  $y(i) - \text{mean}(y)$ , where  $y$  represents the anterior-posterior coordinate in mm of each voxel indexed by  $i$ . A signal-independent orthogonal spatial basis was used in the data-reduction step to produce a set of 7 eigenvariates (1 constant, and 2 low frequency components for each spatial dimension). Results were analyzed in terms of the obtained p-values for the F and T statistics for each region across subjects under a fixed-effect framework.

To compare methodologies we used the corrected voxel- and cluster-level p-values from the SPM analysis, and the F-test p-values from the ROI analysis. In order to appropriately compare the results using both methodologies we computed separately the threshold of the resulting p-values that would equally produce a 5% false-positive rate globally in the right hemisphere following each methodology. Each method's *power* was then defined as the probability of finding a significant activation (resulting p-values below the computed threshold) anywhere in the left hemisphere. Under the null hypothesis (if the signal is undetectable following a given methodology) this measure would approximate the 5% prespecified false-positive rate. Each method's *spatial specificity* was computed from the cases where significant left hemisphere activation was found, as the percentage of times left hemisphere Heschl's gyrus activation was found at a significance level below the prespecified 5% false-positive rate. The *power* measure addresses the question of what is the probability of finding an effect for a given effect size and analysis methodology. The *spatial sensitivity* measure addresses the question of what is, assuming that an effect was found, the probability of correctly labeling its location for a given effect size and analysis methodology.

Figure 5 shows the comparative results for both a standard methodology using inter-subject full-brain coregistration followed by cluster-level or voxel-level analyses and the proposed ROI methodology. In terms of the test power, the proposed methodology (labeled *ROI-level*) compares favorably to the methodologies based on inter-subject full-brain coregistration. ROI-level analyses provide sensitivities to task-related effects of 26%, 98%, and 100% for small medium and large effect sizes respectively. For the same signals, when performing inter-subject coregistration the obtained sensitivities would be 7%, 13% and 88% for voxel-level analyses and

10%, 20%, and 93% for cluster-level analyses. These sensitivities reflect the probability of finding a task-related effect in the left-hemisphere (the actual signal was located in the left-hemisphere Heschl's gyrus) when controlling the false-positive rate in the right-hemisphere at a 5% false-positive level. For small and medium effect sizes the ROI-level analyses provide an overall sensitivity to task-related effects ranging from three to six times the sensitivity obtained using standard voxel- or cluster-level analyses. In terms of the minimal signal size detectable by any methodology at a reasonable power level (>90%), ROI-level analyses are able to adequately identify the presence of a task-related signal with less than half the effect size needed if performing a voxel- or cluster-level analysis.

In terms of the ability of each methodology to correctly locate the source of the signal, standard methodologies based on inter-subject coregistration suffer a considerable degree of ambiguity in the presence of anatomical variability. While the ROI-level analyses provide 100% localizability for all effect sizes tested (when a task-related effect is found in the left hemisphere, it is always present in the left Heschl's gyrus), other methodologies based on inter-subject coregistration provide in the best case (for medium and large effect sizes) between 83% and 85% localizability. For small effect sizes the ability to localize drops abruptly for voxel- and cluster-level analyses (to 35% and 25%, respectively) since the tests' power approximates the pre-specified false-positive rate.

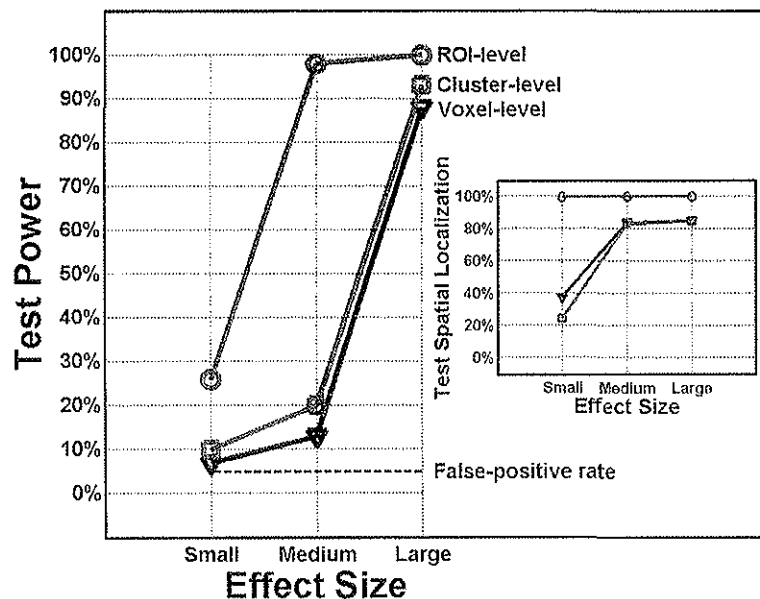
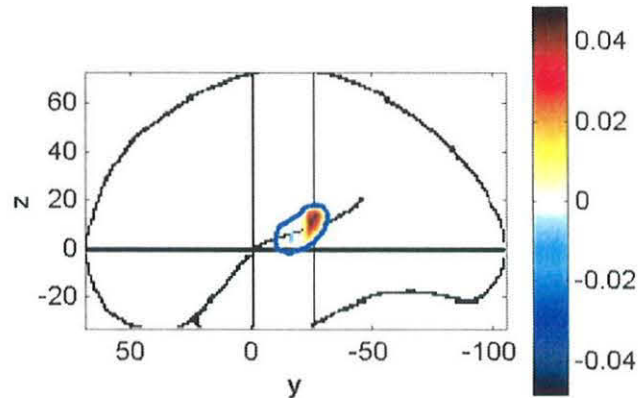


FIGURE 5. Comparative effect of anatomical variability on the power (main plot) and the ability to localize the source of activation (right plot) of three different functional analysis methodologies. Voxel-level and cluster-level analyses are performed following inter-subject full-brain coregistration. ROI-level analyses are performed using multivariate regional analyses on subject-specific ROIs encompassing four times the actual activation loci. Results are shown at three levels of the simulated task-related effects representing signal-to-noise ratios of 1%, 5%, and 10% RMS at the voxel level, respectively. The activation locus was located in the left-hemisphere Heschl gyrus. False-positives are controlled at a 5% hemispheric level (i.e. there is a 5% probability of falsely detecting an activation in the right hemisphere).

The results for the T-test ROI-level analyses (not shown in Figure 5) are similar to the ones obtained for the F-test. The T-test power is 20%, 100%, and 100% for the small, medium, and large effect sizes, respectively, and the corresponding test-localizabilities are 90%, 100%,

and 100%. These results correspond to the tested hypothesis of a spatial pattern of activation showing a differential strength along the anterior-posterior axis within a region (the simulated activation was located in the posterior portion of left HG). The spatial display of the effects on left HG (across-subjects average spatial profile of activation aligning the region centers for each subject) for a medium effect size simulation is shown in Figure 6 on a sagittal “glass” view. Displaying the average spatial profile of activation correctly indicates the activation loci to be located in the most posterior part of this region.



**FIGURE 6. Spatial display of the average response across subjects within left HG for a medium effect size simulations. The thick line represents the average regional boundaries. The activation locus is correctly indicated to lie in the most posterior part of this region. A test on a differential activation within this region along the anterior-posterior axes results in an (hemispheric-level,  $p < .05$ ) significant response.**

Inter-subject anatomical variability has two related effects on the sensitivity of subsequent functional analysis. First, the presence of anatomical variability decreases the overall sensitivity of statistical analyses that are based on whole brain inter-subject coregistration. This is most apparent for small and medium effect sizes for which an ROI-level analysis provides as much as 6 times the sensitivity than other methodologies based on whole brain inter-subject coregistration. Second, the loss of clear regional boundaries increases the false-positive rate above the prescribed value in regions neighboring an area responding to the experimental design, introducing some ambiguity in the regional identity of the observed activation. This is apparent in Figure 5 from the localizability measures saturating at a roughly 85% level for the voxel- and cluster-level analyses. This loss of clear boundaries is induced by the anatomical variability and it is further extended by the spatial smoothing pre-processing stage. Overall the proposed ROI methodology, by combining clearly defined regional boundaries with a multivariate region-level test for the functional activation within these boundaries, provides correct localization and increased sensitivity to experimental task-related effects, and it is also able to appropriately characterize a prominent feature (differential antero-posterior activation) of the within-region spatial profile of activation. The improved sensitivity measures estimated for the ROI methodology are partly based on the correct compensation of inter-subject anatomical variability, which relies on the correctness of the regional definition as the source of functional activation for each subject. In the simulations the defined ROI encompasses four times the size of the actual activation site for each subject. The observed increased sensitivity measures are also partly based on the nature of the region-level analysis, as it provides “pooled” statistics across all the ROI voxels compared to other methods that rely on single voxel statistics. While performing a

standard full-brain normalization followed by an omnibus test limited to an ROI would also provide “pooled” statistics across the ROI voxels and possibly increased sensitivity measures, defining subject-specific ROIs is expected to improve sensitivity by appropriately compensating for inter-subject anatomical variability. Furthermore, omnibus tests relying on Gaussian field theory, as discussed above, do not remain valid for small ROI sizes. Similar to an omnibus test, the proposed technique does not assume spatially focal activation and is expected to provide increased sensitivity in the case of spatially distributed activation responses.

## 5 VALIDATION

The proposed statistical analyses were validated using Monte Carlo simulations with varying noise sources. Unless otherwise stated, the simulated data consists of a single run with 128 scans (TR=2s) over an ROI of 8x8x8 voxels (voxel-size=3mm). A *standard* set of noise model parameters was defined as follows. Low-frequency noise autocorrelation width was set to 25s (FWHM), and wide-band noise was added at each frequency with one-seventh the peak energy of the low-frequency noise (peak ratio 7:1). The noise was further spatially smoothed using a 3mm FWHM Gaussian kernel. The statistical analysis centered on a 1/64 to 1/4 Hz frequency window which represents the full range of frequencies below the Nyquist rate available in the data. False-positive rate measures for the implemented analyses were obtained by running 500 simulations for each condition tested. Each simulation consisted of a region-level T- and F-test on a sinusoidal predictor (T=16s) using the statistical analysis techniques described above. The T-tests used a spatial contrast vector with one in all its elements (testing the presence of an effect on all voxels within a region). Sensitivity measures for the implemented analyses were obtained by again running 500 simulations for each condition tested when adding a sinusoidal signal (T=16s) to all of the voxels with a 1% RMS signal-to-noise ratio.

Seven series of simulations were run. The first one used a standard noise source as defined above. The rest addressed the effect on the tests’ validity and sensitivity of: 1) physiological noise (low-frequency) with extended temporal correlation (60s FWHM); 2) physiological noise with limited temporal correlation (6s FWHM); 3) a small amount of thermal (wideband) noise (2:1 peak ratio); 4) a large amount of thermal noise (20:1 peak ratio); 5) highly spatially-correlated noise (10mm FWHM); and 6) highly spatially-correlated physiological noise (10mm FWHM) with standard thermal noise (3mm FWHM).

For each series of simulations the receiver operating characteristics curve (ROC) was constructed for the case in which no signal was present (addressing the test validity) and for the case in which a signal was present (addressing the test sensitivity).

Figure 7 shows the ROCs (columns 3 and 4) for each of the noise conditions tested (column 1). Thin lines represent the case of no signal present. If the tests are valid they should approximate a diagonal straight line, reflecting that the observed sensitivity (probability of finding a significant activation) is appropriately controlled by the pre-specified false-positive rate. All the obtained ROCs are roughly diagonal. We performed an additional test to confirm that any observed departure in the estimated ROCs from purely diagonal curves reflected variability induced by the limited number of simulations and not a true departure from valid tests. In order to do so we computed 1000 simulated ROCs obtained by extracting 500 samples (the number of simulations run for obtaining the original ROCs) from a true F and T distribution with the correct degrees of freedom. For each of the 1000 newly obtained ROCs we performed a 2-sided Kolmogorov-Smirnov (KS) test on the uniformity of the ROC and computed the 5%

percentile of the obtained KS p values. Last we tested the ROCs obtained from the different noise conditions for uniformity equally using a KS test at the obtained 5% percentile level. None of the ROCs for the different noise conditions rejected the null-hypotheses of uniformity, reflecting that for each of the tested conditions any observed departure from purely diagonal ROCs is within the expected variability given the number of simulations run.

Thick lines represent the case of a signal being present. For all conditions, the signal was consistently detected above the false-positive rate. The T-tests provide in general increased sensitivity, reflecting that they are testing a more precise hypothesis (and the simulated signal has been distributed over all of the ROI voxels). As expected the sensitivity primarily depends on the ratio of the signal energy to the level of noise in the frequency region of the signal, and it is minimal for the condition of physiological noise with limited temporal correlation (since the signal is most masked in the underlying physiological noise extending to the signal frequency).

The combined results in this section show that the proposed tests remain valid for a wide range of possible types of fMRI noise. The standard noise source condition models what we found to be a prototypical noise form using EPI at 1.5T with a repetition time of 2 seconds. The condition of physiological noise with extended/limited temporal correlation is relevant to that expected for faster/slower repetition times. Overall these different noise sources are expected to accommodate differences in scanning conditions. While the validity results shown are obtained without band-pass filtering the temporal series, band-pass filtering will increase the robustness of the statistical tests against noise model misspecification as discussed above.

## 6 CONCLUSION

We have presented a novel methodology for the analysis of functional MRI combining subject-specific region of interest (ROI) definition with a region-level multivariate statistical analysis technique. The proposed ROI-level analyses avoid the need to perform inter-subject full-brain coregistration and spatial smoothing of the functional series. They allow testing functional hypotheses using a general linear model framework regarding the overall activation of specific regions of interest, and hypotheses regarding the spatial profile of activation within those regions. All the techniques necessary to apply the proposed methodology have been implemented in Matlab interfacing with the SPM package and are publicly available (<http://cns.bu.edu/~speech/>). We showed evidence that after a standard full-brain normalization procedure there still exists a considerable degree of residual variability in the shape and location of regions defined based on anatomical markers. Given this observed degree of anatomical variability, we then showed by simulation that the expected sensitivity of standard voxel- or cluster-level functional analyses when based on full-brain normalization techniques is considerably lower than that obtained using the proposed methodology based on region-level analyses. Last, the statistical methods presented were validated using Monte Carlo simulations.

The presented functional analysis is meant as a confirmatory technique, i.e. to test specific model driven hypotheses about the functional response of specific brain regions. By allowing researchers to define and test hypotheses on specific brain regions, replicability and a knowledge buildup from functional results are facilitated. Several aspects of the proposed methodology call for further investigation. Conceptually, better understanding the relationship between functional and anatomical markers is needed. Methodologically, comparisons of the statistical power and validity under different fMRI noise source models and data reduction schemes would help in obtaining more sensible statistical tests. The time-consuming nature of manual subject-specific ROI definition represents a tradeoff for the advantages (better spatial

specificity and improved statistical power) of the proposed methodology. Work on automatic parcellation schemes (Dale et al. 1999) would allow direct implementation of this methodology with greatly reduced effort.

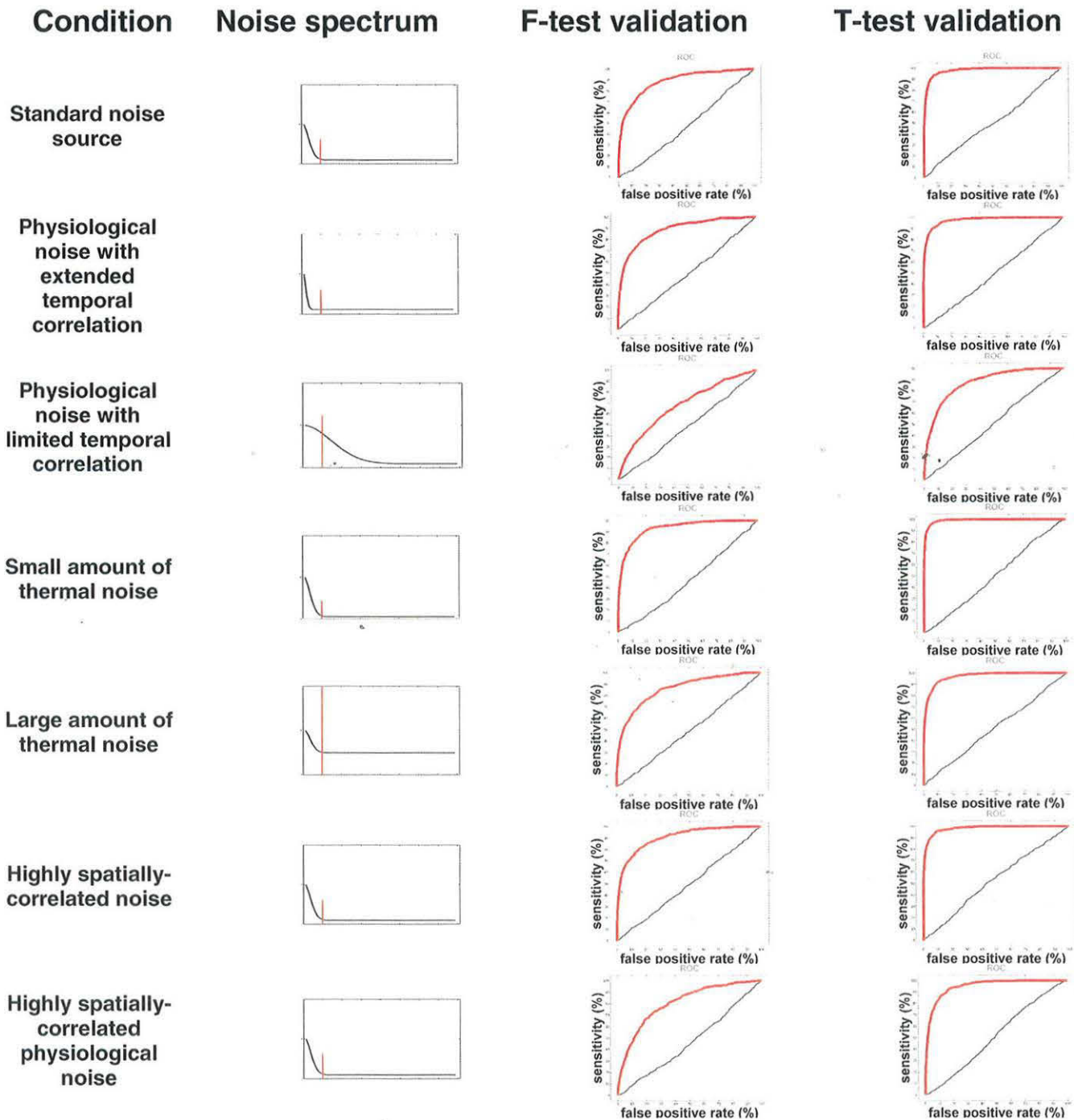


FIGURE 7. Statistical analysis validation results for a variety of possible fMRI noise conditions (column 1). Column 2 shows a schematic of the tested noise frequency spectrum. The simulated sinusoidal activation frequency spectrum appears as a vertical line in these plots. Columns 3 and 4 show the receiver operating characteristic curve (ROC) of the proposed region-level F and T tests, respectively. The test sensitivity in the condition of no signal present (thin line) and signal present (thick line) represent the estimated test false-positive rate and test power, respectively, at different prespecified false-positive levels.

## ACKNOWLEDGEMENTS

We would like to thank Dr. Julie Goodman, George Papadimitriou, and Dr. David Kennedy for their help, and the MGH NMR Center for the use of their facilities. The study was supported by NIH grants R29 DC02852 and R01 DC02852 (Frank Guenther, PI).

## REFERENCES

- Bullmore, E., M. Brammer, S. C. Williams, S. Rabe-Hesketh, N. Janot, A. David, J. Mellers, R. Howard and P. Sham (1996). "Statistical methods of estimation and inference for functional MR image analysis." Magn Reson Med 35(2): 261-77.
- Caviness, V. S., Jr., J. Meyer, N. Makris and D. Kennedy (1996). "MRI-based topographic parcellation of human neocortex: an anatomically specified method with estimate of reliability." J Cog Neurosci 8(6): 566-587.
- Dale, A. M., B. Fischl and M. I. Sereno (1999). "Cortical surface-based analysis. I. Segmentation and surface reconstruction." Neuroimage 9(2): 179-94.
- Friston, K. J., C. D. Frith, R. S. Frackowiak and R. Turner (1995). "Characterizing dynamic brain responses with fMRI: a multivariate approach." Neuroimage 2(2): 166-72.
- Friston, K. J., A. Holmes, J. B. Poline, C. J. Price and C. D. Frith (1996). "Detecting activations in PET and fMRI: levels of inference and power." Neuroimage 4(3 Pt 1): 223-35.
- Friston, K. J., O. Josephs, E. Zarahn, A. P. Holmes, S. Rouquette and J. Poline (2000). "To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis." Neuroimage 12(2): 196-208.
- Mardia, K. V., J. T. Kent and J. M. Bibby (1979). Multivariate analysis. London ; New York, Academic Press.
- Purdon, P. L. and R. M. Weisskoff (1998). "Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI." Hum Brain Mapp 6(4): 239-49.
- Rademacher, J., V. S. Caviness, Jr., H. Steinmetz and A. M. Galaburda (1993). "Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology." Cereb Cortex 3(4): 313-29.
- Worsley, K. J., C. H. Liao, J. Aston, V. Petre, G. H. Duncan, F. Morales and A. C. Evans (2002). "A general statistical analysis for fMRI data." Neuroimage 15(1): 1-15.
- Worsley, K. J., J. B. Poline, K. J. Friston and A. C. Evans (1997). "Characterizing the response of PET and fMRI data using multivariate linear models." Neuroimage 6(4): 305-19.
- Zarahn, E., G. K. Aguirre and M. D'Esposito (1997). "Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions." Neuroimage 5(3): 179-97.