Group Analysis of Functional Imaging Data using

Penalized Maximum Likelihood

Rao P. Gullapalli\textsuperscript{1}, Ranjan Maitra\textsuperscript{2}, Steven R. Roys\textsuperscript{1}, Joel Greenspan\textsuperscript{3},
Gerald V. Smith\textsuperscript{4}, Gad Alon\textsuperscript{4}

\textsuperscript{1}Department of Radiology, University of Maryland School of Medicine, Baltimore, MD 21201
\textsuperscript{2}Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD 21250
\textsuperscript{3}Department of Oral and Cranial Biological Sciences, University of Maryland Dental School, Baltimore, MD 21201
\textsuperscript{4}Departments of Physical Therapy, University of Maryland School of Medicine, Baltimore, MD 21201

Total Word Count: 5009

Corresponding Author:

Rao P. Gullapalli
Department of Radiology
22 S. Greene St
University of Maryland, Baltimore
Baltimore, MD 21201
Phone: 410-328-2099
Fax: 410-328-0341
e-mail: rgullapalli@umm.edu

Running Title: Group Analysis of fMRI data
The value of the penalized maximum likelihood method was recently shown in assessing the test-retest reliability of functional activation (Maitra et al: MRM 2002; 48:62-70). We extend this methodology to the analysis of grouped functional magnetic resonance imaging (fMRI) data. Specifically we have applied this technique to two functional paradigms, (a) pain paradigm that used a mechanical probe to provide noxious stimuli to the dorsum of the left foot, and (b) four levels of graded peripheral neuromuscular electrical stimulation (NMES). Reliability of activation maps were generated for both the paradigms. Receiver operator characteristic (ROC) curves were generated in the case of the graded NMES paradigm for each level of stimulation. These curves revealed an increase in the specificity of activation with increasing stimulus levels. Further a methodology was developed using the maximum likelihood method to determine whether the grouped reliability maps obtained from one stimulus level were significantly different from the previous levels. Our results show a significant difference (p<0.01) in reliability of activation from one stimulation level to the next. These results are in agreement with the results obtained using voxel-by-voxel measures of functional MRI signal intensities and spatial extent of activation. Besides providing information on the performance of the paradigm in a group, this methodology can also be used to optimize novel paradigms with a goal of minimizing the false detection rate.

**Keywords:** fMRI, quantitation, Markov Random Field, Iterated Conditional Modes, ROC analysis, somatosensory, neuromuscular electrical stimulation
Introduction

Cognitive neuroscientists and clinicians over the past decade have been investigating several novel functional magnetic resonance imaging (fMRI) paradigms and acquisition techniques in order to obtain insight into normal and abnormal brain function. However, the analysis of such data still presents formidable challenges. This problem is compounded by the fact that both inter- and intra-subject variations are observed when using the same paradigm. Inter-subject variations pose an even bigger challenge to the objective analysis of grouped data especially when analyzing cortical responses to novel paradigms. Recently we presented a comprehensive approach to address the question of test-retest reliability using fMRI data on individual subjects.1 This method was based on data from repeat scans on the same subject using the same paradigm and used a penalized maximum likelihood criterion that incorporated both temporal and spatial information.1-3 This method estimated the mixing proportion $\lambda$ on a voxel-by-voxel basis from M replications of a study, using the probabilities of a voxel being correctly identified as active ($\pi_A$) and of a voxel incorrectly identified as inactive ($\pi_I$). A reliability map was then produced showing cortical areas of the brain having a high reliability of activation. A framework was also provided for determining the optimal threshold through the use of the maximum likelihood (ML) reliability efficient frontier. The ML reliability efficient frontier was defined for each threshold as the probability that the true state of the voxel, whether active or inactive, is correctly identified for that threshold. This methodology was indifferent to the basic processing methods employed in that the input to this methodology could use functional maps generated by any of several methods such as cross-correlation, t-test, etc. In this paper we provide a framework for the applicability of this methodology to analyze grouped data. We show the application of this methodology on two different paradigms, (1) a pain paradigm
involving noxious mechanical stimulation, and (2) a graded neuromuscular electrical peripheral nerve stimulation (NMES) paradigm to assess sensory and motor responses. The methodology and applicability of the proposed method to reliably differentiate activation elicited by different levels of stimulation is also described in the case of the electrical peripheral nerve stimulation paradigm.

**Methods**

**Imaging**

All MR images were obtained on a GE 1.5 Tesla Signa system using v5.8 software and equipped with echo-speed gradients. The imaging protocol consisted of the following scans: (1) A T1-weighted structural scan with TR/TE of 500/12 ms; (2) a series of functional MRI scans using single shot spiral imaging technique\(^4\) at a TE of 35ms and a TR of 3000ms for the pain paradigm, and a TR of 4000ms for the NMES paradigm; (3) a phase contrast angiography scan covering the same locations as the functional scan to depict major draining veins; (4) and a volumetric scan covering the entire brain at a TR/TE/flip of 25ms/4.6ms/20\(^\circ\) respectively. For the noxious pain paradigm, 16 slices were obtained at a thickness of 6mm with no gap. For the peripheral electrical stimulation scans the whole brain was covered using 24 slices at 6mm each with no gap.\(^5\) Six and ten consenting, neurologically intact subjects were scanned using the pain paradigm and the NMES paradigm respectively. For the pain paradigm, a mechanical probe with a weight of 60gms and a tip size of 0.01 mm\(^2\) was used to provide the noxious stimulus to the dorsum of the left foot. This level of stimulation has been shown to be above pain threshold psychophysically and above nociceptor activation threshold electrophysiologically.\(^6,7\) Padding around the head was provided along with taping the forehead to the headrest to minimize subject
motion during the paradigm.\textsuperscript{6} The paradigm consisted of eight cycles of 15s on and 15s off boxcar type stimulus for a total scan time of 4 minutes. Additional precaution was taken for minimizing motion during the execution of the NMES paradigm since the large movement of the limbs could induce severe motion. Subjects assumed a supine position on the gantry table of the scanner and the head, torso, and lower limbs were stabilized and supported by a custom-designed torso harness and knee bolster system, to prevent head motion during scanning.\textsuperscript{8} Two self-adhesive (7.6 x 12.7 cm) electrodes were secured over the right quadriceps muscle and connected to shielded leads that extended out of the MRI room and were connected to a portable neuromuscular stimulator and digital oscilloscope. Four stimulation intensity levels:

1) Sensory threshold [Th],
2) (Max-Th) *0.333 +Th [low-Intermediate]
3) (Max-Th)*0.666+Th [high-intermediate], and
4) motor stimulation resulting in full knee extension [Max]

were used to dose the amount of peripheral excitation. Stimulus intensities were quantified by stimulus phase charge and peak current.\textsuperscript{9} Boxcar stimulation protocol that included 8 cycles of 20 sec stimulation and 20 sec relaxation was used for a total scan time of 5 min 20 seconds per stimulus level. The sequence of introducing the graded stimulus levels was randomized.

Data was transferred to an SGI Origin 200 workstation where reconstruction of the functional data, time series generation, motion correction, and pixel normalization was performed using local scripts generated to run with the package Analysis of Functional Neuroimages (AFNI) developed and freely distributed by Dr. Robert Cox.\textsuperscript{10} Time series correlation images were then generated using a boxcar reference function. After correlation,
images of all the subjects along with their functional overlays were transformed to a common Talairach coordinate space and resampled to 2mm cubical voxels for group analysis.\textsuperscript{11}

To alleviate concerns over the magnetic field interaction of the electrical peripheral nerve stimulation with the resulting activation, in addition to the above scans, one subject within this group was scanned twice. During the latter scan, a large (12.7x25.4 cm) carbon-silicate electrode barrier was placed between the stimulating electrodes and the subject's skin. The subject did not receive the stimulation current although a complete circuit was made and the standard scanning protocol was followed for each stimulation level. Subsequent analysis of this "sham stimulation" data showed no activation in any of the areas of interest and at any of the stimulation levels. Additionally, the scan demonstrated that the electrical currents generated by the stimulator separated from the brain sufficiently to preclude any measurable alterations or distortions of the magnetic field in the brain regions.

Statistical Methodology

The processing of the data has been discussed elsewhere and Fig.1 shows the processes involved in the generation of the reliability maps. Briefly, the processing uses both spatial and temporal information of the fMRI data to provide a probability map of true activation.\textsuperscript{1} Scans from various subjects at a number of correlation thresholds were processed using the Iterated Conditional Modes (ICM) implementation for the method of penalized maximum likelihood.\textsuperscript{12} This methodology was used to estimate probabilities of both true positives and false positives ($\pi_A$ and $\pi_I$, respectively) as well as the voxel-specific probabilities of a voxel being truly active.
These estimates were combined to obtain the reliability and anti-reliability measures. The reliability of an actively identified voxel at a given correlation threshold was defined as the probability of a voxel being truly active, when it has been identified as active at that threshold. The anti-reliability measure of an inactively identified voxel at a correlation threshold was defined as the probability of it being truly active at a given threshold, when it has been identified as inactive at that threshold. The penalty function used in the procedure incorporates spatial context in the estimation of the map representing the probabilities of the activation and was added as a constraint to the logarithm of the likelihood function. A Markov Random Field prior distribution was used on the map of probabilities of true activation to characterize this spatial context. Further analysis of the data was also done using the receiver operator characteristics (ROC) curves in the case of the multiple NMES to test for the significance of one stimulation level to the next. The ROC curves provided a measure of the false detection rate (false positives) at any given threshold of activation.

After generating reliability and anti-reliability maps for both the somatosensory and the NEMS paradigms, we posed the following question to the data derived from the NMES stimulation paradigm: Are there significant differences in test-retest reliability estimates obtained at different stimulation intensity levels? To answer this question, we formulated it in the context of a hypothesis testing problem where the null hypothesis to be tested is whether the true values of the voxel-specific $\lambda$'s and the threshold level-specific $\pi_A$'s and $\pi_I$'s are the same over all stimulation levels, against the alternative that they are not all the same. Formally, the null hypothesis can be stated as

$$H_0: (\lambda_m, \pi_{Am}, \pi_{Im}) = (\lambda_j, \pi_{Aj}, \pi_{Ij}) \text{ for all } 1 \leq m < j \leq 4$$
against the alternative

\[ H_A: (\lambda_m, \pi_{Am}, \pi_{Im}) \neq (\lambda_j, \pi_{Aj}, \pi_{Ij}) \text{ for all } 1 \leq m < j \leq 4. \]

We used the likelihood ratio test for this problem. To describe the test statistic, we extend notation used in Maitra et al.\(^1\) Consider the following setup: Let \( \lambda_j \) be the probability that the \( r \)th voxel is truly active at the \( j \)th stimulus level. Note that there are four such stimulus levels.

Let \( N \) be the total number of voxels in the image cube and let \( K \) be the number of activation threshold levels. Let \( Y_{ji} = (y_{j,i,1}, y_{j,i,2}, \ldots, y_{j,i,K}) \), where \( y_{j,i,k} \) is the number of replications for which the \( r \)th voxel is active at the \( k \)th activation threshold and at the \( j \)th stimulus level.

Without loss of generality, assume that the threshold levels are in increasing order. Further, let \( p_{Aj,k,k-1} \) be the probability at the \( j \)th stimulus level of a truly active voxel being so identified at the \( k \)th threshold level given that it is also correctly identified as active at the \((k-1)\)th threshold level. Similarly, let \( p_{Ij,k,k-1} \) be the corresponding probability at the \( j \)th stimulus level of a truly inactive voxel being identified as active at the \( k \)th threshold level, given that it is also incorrectly identified as active at the \((k-1)\)th threshold level. These relate to \( \pi_{Aj,k} \) and \( \pi_{Ij,k} \) as

\[
\pi_{Aj,k} = P(\text{true active voxel is identified as active at the } k \text{ th threshold and the } j \text{ th stimulus level})
\]

\[
= \prod_{l=1}^{k} P(\text{true active voxel is identified as active at the } l \text{ th threshold and } j \text{ th stimulus level} \mid \text{it is also so identified at } (l-1) \text{ th threshold and } j \text{ th stimulus level}) = \prod_{l=1}^{k} p_{Aj,l,j-1}
\]

and

\[
\pi_{Ij,k} = \prod_{l=1}^{k} p_{Ij,l,j-1}
\]

For notational consistency, we denote \( p_{Aj,1,0} \) and \( p_{Ij,1,0} \) as \( \pi_{Aj,1} \) and \( \pi_{Ij,1} \), respectively, which are the corresponding probabilities at the \( j \)th stimulus level of a truly active voxel being identified as active at the first threshold level. Also, let \( y_{i,j,0} = M \) be the total number of

8
subjects. Finally, let $\lambda_j$ be the vector of voxel-specific $\lambda$-values for the $j$'th stimulation level and $\pi_{Aj}$ and $\pi_{Ij}$ be correspondingly, the vectors of values of the $\pi_{Ajk}$'s and $\pi_{Ijk}$'s for the $j$'th threshold level. Then the likelihood-ratio-test statistic is given by ‘log($\Lambda$)’ where

$$\Lambda = \frac{\prod_{j=1}^{n} \prod_{i=1}^{n} D(\hat{\lambda}_{0i}, \pi_{A0i}, \pi_{I0i})}{\prod_{j=1}^{n} \prod_{i=1}^{n} D(\hat{\lambda}_{ji}, \pi_{Aj}, \pi_{Ij})}$$

[3]

Here the numerator is the likelihood evaluated under the assumption that the parameter values of $(\lambda_{i}, \pi_{A_{ij}}, \pi_{I_{ij}})$ is the same for all levels and

$$D(\lambda_{*j}, \pi_{A_{*j}}, \pi_{I_{*j}}) = \lambda_{j} \prod_{k=1}^{K} \left( \frac{y_{i,k-1}}{y_{i,k}} \right) p_{Ak, k-1}^{y_{i,k-1}} (1 - p_{Ak, k-1})^{y_{i,k} - y_{i,k-1}} + (1 - \lambda_{j}) \prod_{k=1}^{K} \left( \frac{y_{i,k-1}}{y_{i,k}} \right) p_{Ik, k-1}^{y_{i,k-1}} (1 - p_{Ik, k-1})^{y_{i,k} - y_{i,k-1}}$$

[4]

Under this assumption, the joint likelihood of the observations over all stimulation levels are just four independent copies of the likelihood at any level and is therefore given by the product in the numerator at the maximizing values of the common $(\lambda, \pi_{A_{i}}, \pi_{I_{i}})$'s. These values and the maxima are calculated using the ICM and penalty function as in Maitra et al.\textsuperscript{1,12} For the denominator, the joint likelihood is the product of the maximized likelihood's at the individual stimulus levels -- each also obtained using ICM and the penalty function of Maitra et al.\textsuperscript{1} This enabled the computation of the likelihood ratio test statistic above.

The $p$-value of the test statistic cannot be obtained analytically. We therefore used simulation and the parametric bootstrap to estimate the $p$-value.\textsuperscript{17} Simulation data sets were
generated at each stimulus level using the common estimates obtained under the null hypothesis. From each data set, the value of the test statistic was computed as above. The \( p \)-value of the observed test statistic is the proportion of times that it lies below the reference distribution of the test statistics obtained using the simulation data sets. For practical reasons, we used a bootstrap sample of size 100 and used that to estimate the \( p \)-value of the test statistic.

Having tested for difference in reliabilities, we tested for whether the reliabilities and parameters were significantly different for each set of successive stimulus levels. This can once again be cast within the framework of a hypothesis-testing problem. Formally therefore, we tested the null hypothesis

\[
H_0: (\lambda_i, \pi_{A,j}, \pi_{I,j}) = (\lambda_{j+1}, \pi_{A,j+1}, \pi_{I,j+1})
\]

against the alternative

\[
H_A: (\lambda_j, \pi_{A,j}, \pi_{I,j}) \neq (\lambda_{j+1}, \pi_{A,j+1}, \pi_{I,j+1}) \text{ for each } j=1,2,3,4 \text{ stimulation levels.}
\]

Once again we performed a likelihood ratio test and computed the test statistic similar to the one above, with the exception that we only use the two successive stimulus levels (instead of all four). We again estimated the \( p \)-value of the test statistic using simulation via the parametric bootstrap.

**Results**

Figure 2 shows the \( \lambda \) maps of the probability of any given pixel being truly active using the above method for the pain paradigm. The opaque yellow regions in the figures have high
reliability, and very little chance of being false positives. Group analysis shows high probability of activation in a region consistent with the second somatosensory cortex bilaterally (more posterior regions of the upper images), and the anterior cingulate cortex (lower images), which agrees with previous pain studies.¹⁸

Figure 3 provides the activation maps at four different stimulus levels in the sensorimotor regions of the cortex for the NMES paradigm. The corresponding ROC curves from the entire data set for the NMES paradigm are shown in Fig. 4. This figure shows the relation between true positive and false positive voxels at each stimulus level. Figure 5 shows a comparison of the ROC curves derived from the whole brain data versus data derived from localized motor region only. ROC curve derived form localized regions provide an improved specificity as they provide a reduction in the false detection rate.

When testing to see whether the voxels specific λ's and the threshold specific πₐ's and πᵢ's were the same at all stimulus levels, the value of the observed test statistic was -4.898x10⁴ (p < 0.01) suggesting a rejection of the null hypothesis. This shows that the test-retest reliability parameter estimates are significantly different at the different stimulus levels. When testing for significant differences in parameter values between the lowest stimulus level and the one higher, the test statistic was evaluated to be -2.472x10⁵ (p<0.01). The test statistic for testing for significant differences at the second and third stimulus levels was evaluated as -9.601x10³ (p<0.01). The parameter values were also significantly different for the third and highest stimulus levels with a test statistic of -2.894x10⁴ (p<0.01). Thus, our results indicate that stimulus levels (dose) significantly affect the reliability of activation (response).
Discussion

Recently we developed a technique that provided test-retest reliability of fMRI data.¹ This technique estimated voxel-specific true activation rates $\lambda$ by taking into consideration the activation state of the neighboring voxels and applying the maximum likelihood method to arrive at a reliability estimate. In that study, using a motor paradigm a method was described to arrive upon a measure for the minimum number of repeat scans necessary in order to arrive at the reliability map. It was determined that a minimum of 5-6 scans from different sessions was necessary for the same subject to account for variability in fMRI data. Further, we also introduced the concept of anti-reliability and the choice of the optimum threshold through the use of the maximum likelihood efficiency frontier. Briefly, anti-reliability was defined as the regions that were determined to be inactive compared to the reliability measures obtained from baseline reliability maps. The ML reliability efficiency frontier provides optimal thresholds using $\lambda$, $\pi_A$, and $\pi_I$ at different activation thresholds for maximizing the true positive rate.

While our objective in that study was to examine the variability of fMRI data within subjects to arrive at the reliability maps, similar questions are of interest when embarking upon a novel paradigm to examine its effects on a group of subjects.

In this study we looked at the applicability of the method to group analysis of noxious pain and peripheral nerve stimulation data. Noxious pain data derived from mechanical stimulation clearly revealed focal activation in the region of S2 (i.e., inferior, anterior parietal lobe), and the anterior cingulate cortex as shown in Fig. 2. Group analysis also indicates bilateral activation in the S2 area, albeit weak on the ipsilateral side to the stimulus applied. We did not detect any significant activity in the S1 area using this paradigm. These results are in agreement with
previously reported pain imaging studies, in which S2 and anterior cingulate gyrus activation is routinely detected, but S1 activity is more tenuous.\textsuperscript{18,19}

While our analysis using the noxious pain paradigm clearly shows that group analysis of fMRI data allows one to make inferences about the possible involvement of different cortical areas, we were also interested in the applicability of our technique to elucidate a dose-response relationship. The dose can be in the form of the strength of the stimulus applied, or the frequency at which it is applied. The response is measured in the relevant cortex where the BOLD activity presumably scales with the intensity of the stimulus applied. Since the BOLD activity can vary significantly both within and across subjects performing the same paradigm, we applied our group analysis technique to a dose-response paradigm to evaluate whether the differences in response from different stimulation level could be elucidated. Figure 3 clearly shows a dose-response relationship between NMES and brain activity in the sensorimotor area. Similar responses were also seen in the cerebellum, cingulate gyrus, and the thalamus. This response is also evidenced by the ROC curves shown in Fig. 4, which clearly shows that at any given threshold the false detection rate is reduced as the stimulation intensity is increased. The technique developed here assesses the reliability of a voxel identified as activated across subjects. In this methodology we calculate the probabilities of both true positives (a truly activated voxel being so identified, type II error) and false positives (an inactivated voxel being misclassified as active, type I error). As can be seen from the ROC of Fig. 4, the proportion of true positive voxels increases with increasing stimulation levels. Using this methodology, the voxel specific threshold maximizing the probability of a voxel being correctly identified as active or inactive (i.e., the sum of the probability of a truly activated voxel being identified as
active and the probability of a truly non-activated voxel identified as not active) is given by the point on the ROC curve with gradient, \((1 - \lambda_{j,i})/\lambda_{j,i}\) where \(\lambda_{j,i}\) is the probability of the \(i'\)th voxel being truly active at the \(j'\)th stimulus level.\(^{20}\) When extracting quantitative information from fMRI data, especially when working with a new paradigm, most researchers are faced with the issue of an appropriate threshold level to use. It is hoped that this new strategy that provides a rational methodology for determining optimal thresholds for subsequent data analysis will facilitate future fMRI studies, particularly those studies that evaluate longitudinal changes or those that look for a dose-response relationship in brain activation. As in the present case, such a strategy is especially important when using data pooled across large numbers of subjects. These results are concordant with voxel by voxel measurement of the signal intensities and spatial extent as shown in Fig. 6 for activation in the sensorimotor areas. This figure shows the group signal intensity response and the spatial extent of activation. Similar pattern of behavior was observed in the cerebellum, cingulate gyrus, and the thalamic regions. While individual subject responses varied considerably, using voxel-by-voxel analysis, high and significant correlations were found between the level of stimulation and amplitude of activation in S1 \((R^2=0.99)\), M1\((R^2=0.88)\), cerebellum \((R^2=0.94)\), thalamus \((R^2=0.80)\) and cingulate gyrus \((R^2=0.91)\). High linear correlations varying between \(R^2=0.91-0.94\) were likewise found between stimulation amplitude and the maximum correlation coefficient \(r\) value of fMRI recordings in these regions. The spatial extent of activation also correlated linearly with the applied stimulus \((R^2=0.86-0.98)\) in these regions. While our results correspond well with this simple approach it is based on a more accurate model, which uses both spatial and temporal context that incorporates the noise characteristics. Further our methodology can be readily automated for routine use thus avoiding the tedious voxel-by-voxel analysis prior to grouping the data.
Two different group analysis techniques are in common use depending upon the type of question the investigator poses. The 'region of interest' or the ROI technique where statistical maps from individuals obtained from user defined ROI's (based on anatomical locations) are grouped together. Such an approach is typically used when investigators have apriori knowledge of the activation region to obtain the statistical differences in these regions between two or more groups. The second approach is typically used when exploring novel paradigms and the neurobiology of the cortical function has not yet been established. Typically in this case, the individual's data is transformed into a common stereotactic atlas such as the Talairach atlas as was done in this study. Statistical inferences then point to the neural circuitry involved within a group, as well as indicate differences between two groups. The methodology described in this study benefit either of the two types of situations described above. Using our method with the ROI approach will provide the investigator a better understanding of the regional variation while also providing a means to investigate the region optimally as shown in Fig 5. The methodology serves well in the case of investigating novel paradigms as a group as we have shown with the NMES study. It allows investigators to identify the major neural circuitry involved and allows them to delineate finer neural processes. While we have shown the applicability of our method in the context of a dose-response scenario, the method can be applicable to the longitudinal study of a group undergoing therapy or to understand the differences between two groups that may differ in their cortical processing.

As discussed in our previous report on the test-retest reliability of fMRI data it should be emphasized that the results of the present study depend totally on the statistical analysis chosen to derive the individual statistical maps. While we have used individual correlation maps as the
basis for the group analysis, the readers should be aware of more sophisticated multivariate analysis tools to generate the functional maps that would be appropriate for their analysis. It should also be noted that the reliability of this technique is limited by the reliability of the transformation of individual brains to a common stereotactic atlas. It has been shown that even after mapping to a common atlas such as the Talairach coordinate systems the variability in the location of individual structures can be as high as 20 mm.\textsuperscript{22} It is possible that such a variation can result in a loss of specificity in localizing the activity within a group.

Applying the method of Maitra \textit{et al.} to this current scenario implies an assumption that variability in the activation due to stimulus level is far more than inter-subject variability.\textsuperscript{1} This means that relative to the variability in activation due to stimulus, the inter-subject variability is less. This is the basis for several recent research papers but has not been tested. A way to test this assertion, outside the scope of this paper, would be to obtain replications for a number of subjects at each stimulus level and to perform the statistical test.

In conclusion we have developed a methodology that provides reliable activation maps from a group of subjects performing a specific paradigm. Further we have shown the applicability of this method in the context of a dose-response paradigm. This methodology also provides a means for choosing the optimal threshold value for a group to maximize the true positives. In addition it provides a framework to bring about the difference in two or more groups subject to different treatments.
ACKNOWLEDGEMENTS

This work was partly supported by a grant from the NIH (R01-NS38493).
REFERENCES:


Figure Captions

Figure 1. Flowchart showing the voxel-by-voxel processing involved in determining the mixing proportion of active voxels $\lambda$; probability of a voxel being correctly identified as active, $p_A$; probability of an inactive voxel falsely identified as active, $p_I$; and, $\beta$ which measures the strength of interactions between neighboring $\lambda$'s.

Figure 2. Lambda images showing brain activations using the 60gm weighted probe applied to the dorsum of the left foot. The functional overlays are at a correlation threshold of 0.65 in the S2, and the anterior cingulate gyrus.

Figure 3. Lambda images showing brain activations in the sensorimotor areas using the peripheral neuromuscular electrical stimulation paradigm at the (a) sensory threshold, (b) low-intermediate threshold (0.33*Max stimulus), (c) intermediate-medium threshold (0.66*Max stimulus), and (d) maximum threshold. The maximum Lambda for each of the stimulus levels was 0.030043, 0.039747, 0.066231, and 0.123879 respectively. Please note the reduction in noise as the stimulus strength is increased with a corresponding increase in the strength of the activation.

Figure 4. Receiver operator characteristics (ROC) curves showing a reduction in false positives ($\pi_I$) for a given true positive ($\pi_A$) with increase in peripheral neuromuscular electrical stimulation. The shift in the curves to the left with increasing stimulus levels indicates the increase in specificity in determining truly active voxels while maintaining high sensitivity.

Figure 5. ROC curves from examining a localized region involving three slices encompassing the motor region showing a further increase in specificity using the peripheral neuromuscular electrical stimulation paradigm at the minimum threshold (sensory).

Figure 6. Bar graphs showing fMRI signal intensity changes and the spatial extent of activation in the sensorimotor cortex at different stimulus intensity levels using the peripheral neuromuscular electrical stimulation paradigm. Numbers 1-4 correspond to the four-graded stimulus levels from minimum to maximum. Voxels that were active in the sensorimotor region at 85% of maximum threshold were included in this graph.
Acquire BOLD time series images

Register the BOLD time series images

Normalize the BOLD time series images

Perform correlation analysis on the normalized BOLD time series images

Transform the images into Talairach space

Compute an initial lambda image

Compute initial $p_A$'s and $p_I$'s for the thresholds

Estimate new $p_A$'s and $p_I$'s from the lambda image and the current $p_A$'s and $p_I$'s

Estimate a new beta from the lambda image and the current beta

Estimate a new lambda image from beta, the $p_A$'s and $p_I$'s, and the current lambda image using a single-step ICM process

Have we converged?

No

Yes

Exit
Figure 2

Lambda image, 60g, 0.65 threshold
\[ \lambda = 0.003411 - 0.02379 \]
Figure 3
Figure 5
Figure 6