Development of an Automated Threshold Technique Based on Reproducibility of fMRI Activation.

T. Stevens[1,2], S. Beyea[1,2], R. D'Arcy[2,3], D. Clarke[4,5], C. Bowen[1,2], and G. Stroink[1]
[1]Physics, Dalhousie University, Halifax, NS, Canada, [2]NRC Institute for Biodiagnostics (Atlantic), Halifax, NS, Canada, [3]Neuroscience, Dalhousie University, Halifax, NS, Canada, [4]Neurosurgery, QEII Health Science Center, Halifax, NS, Canada, [5]Surgery, Dalhousie University, Halifax, NS, Canada

Introduction: Setting the activation threshold remains a challenging problem in functional magnetic resonance imaging (fMRI). Existing strategies for determining threshold levels include voxel-wise confidence levels, multiple comparison corrections and clustering requirements [1,2]. Clustering algorithms have the advantage of eliminating small, isolated activation regions, but may potentially eliminate valid activation foci. Corrections for multiple comparisons address the increased chance of false positive activations due to the large number of voxels in an fMRI image, but typically cannot account for differences in activation strength between tasks, individuals, or scanners [2]. Setting appropriate thresholds is particularly pertinent in presurgical mapping, where knowledge of the location and extent of functionally eloquent cortex can affect surgical decisions regarding approach and completeness of tumor removal. In this work, we demonstrate an automated threshold technique based on test-retest imaging and receiver-operator characteristic (ROC) curves [3,4]. This procedure is capable of producing individualized threshold levels that are optimized for the reproducibility of the observed spatial pattern of activation.

Methods: Anatomical and functional images were collected from patients with cortical brain tumors using a 4T MRI. Each patient was presented with a site-directed task battery chosen based on the location of his or her tumor. The subjects were asked to perform each task twice in a single session to permit the creation of test-retest ROC curves. Functional imaging was performed using a spiral pulse sequence (2-shot, TR 2s, TE 15ms, FOV 24cm, 64x64 matrix, 20-22 slices, 4mm with 0.5mm gap), registered to a T1 weighted anatomical (MP-FLASH, TR 10ms, TI 500ms, α 11°, FOV 24cm, 256x256 matrix, 170-190 slices, 1mm with no gap). The functional images were realigned for motion correction, registered to the anatomical image, smoothed with a 6mm FWHM, and analyzed using the standard generalized linear model approach in AFNI. The resulting t-statistic functional maps were used for subsequent post-processing and optimization.

The ROC curve is a plot of true and false positive rates (fraction of voxels that are correctly and incorrectly identified active to total number of truly active and inactive voxels respectively). The automated ROC optimization procedure uses one of the test-retest images (test) as the activation template to determine what is truly active and inactive. For a given pair of test and retest image threshold, the true and false positives rates are determined using an algorithm that compares each voxel in the test image to the voxel at the same location in the retest image. If both voxels are above (below) threshold, it is declared a true positive (negative), and a false positive (negative) if below (above) threshold in the template but above (below) in the retest. This procedure is repeated for all possible values of the test and retest image threshold, to produce a family of ROC curves.

The area under the ROC curve, representing the average value of the sensitivity, is plotted with its first derivative as a function of the test image threshold. The test image threshold above which further increases in threshold return only marginal gains in area under the retest ROC curve (derivative of 0.1 or less) is selected. From this ROC curve, the retest threshold that produces the greatest combined sensitivity and specificity is determined and applied to the retest image. The procedure is repeated with the roles of each image switched to determine thresholds for both images.

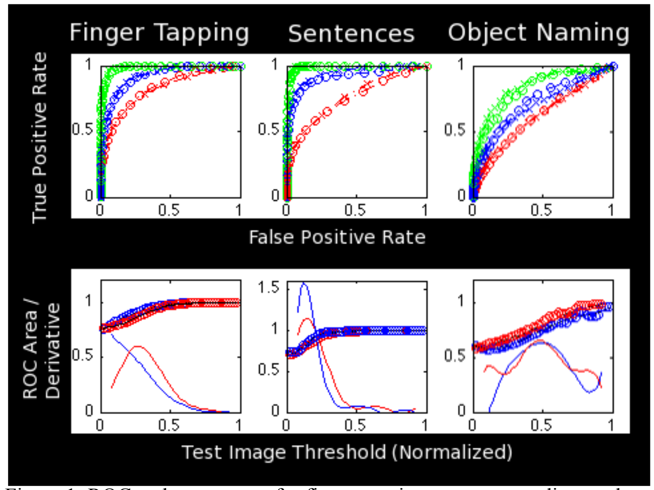

Figure 1: ROC and area curves for finger tapping, sentence reading, and object naming tasks. ROC curves correspond to 1% of voxels active in the template (green), 10% (blue) and 50% (red), for both the test (circles) and retest (exes) images as template. Area versus threshold curves are shown for the test image as template (red circles) and retest as template (blue circles). The derivative is shown on the same graph as a solid line.

Results: The ROC based threshold procedure described above has been applied to mapping of hand and foot motor systems, and to investigations of function of the frontal and temporal language zones. Figure 1 shows the ROC curves produced for a selection of representative tasks. ROC curves are shown for both the test and retest image acting as the template, and for three values of the template threshold, corresponding to 1, 10, and 50 percent of the voxels in the template above threshold. In 5 of 8 test-retest datasets analyzed, the area under the curve is 0.95 or greater for the 1 percent active template. Figure 2 shows the activation maps resulting from applying the ROC determined threshold and the threshold corresponding to an FDR multiple comparison correction of q=0.01.

Discussion & Conclusion: The area under the ROC curve for all motor tasks exhibits predictable behavior, with an initial region of high first derivative as the threshold is increased, reaching a gradual plateau, at which point the derivative drops dramatically. Cognitive tasks are more variable, and often have more gradual increases in area under the ROC curve, although this is not always the case as can be seen by comparing the sentence comprehension and object naming ROC curves in figure 1. The object-naming task involves overt language generation, and the reduced quality of the ROC curves is likely a result of task correlated motion artifact. As the derivative of the area for this task does not reach the specified cut-off, the retest image threshold for object naming is set from the ROC curve with the largest area obtained. The image thresholds determined by the algorithm developed in this work (t=4.0-7.8; figure 2 top) are typically more strict than those obtained by the FDR method (t=2.9-3.4; figure 2 bottom). Despite the poorer performance of the object-naming images during the optimization routine, the extracted pattern of activation included much of the putative frontal language zones.

The goal of presurgical mapping is to determine the most important brain structures to respect during surgery. The automated ROC approach to setting fMRI thresholds could be valuable in this context, as the reduced extent and number of activated regions are simultaneously assured to have a reasonable level of reproducibility. As our sample is drawn from the brain tumor population, we are able to collect cortical stimulation data from the same series of patients, which has also allowed us to investigate the diagnostic utility of the ROC threshold procedure. Importantly, this automated procedure for setting image thresholds is insensitive to differences in overall magnitude of activation, and relies only on the underlying spatial distribution of activity to be the same in both scans, an assumption that is at the core of diagnostic neuroimaging.

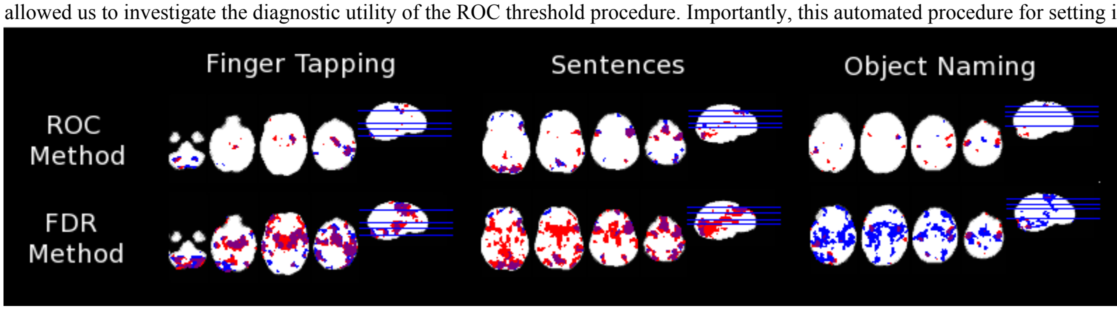

Figure 2: A selection of slices depicting the test (red) and retest (blue) images with automated ROC threshold applied (top) and manual FDR threshold (q=0.01, bottom). Overlap of the two images is shown in purple. All images are in radiological view.

References: [1] B. Logan and D. Rowe, *NeuroImage,* 22: 95-108 (2004); [2] R. Heller et al., *NeuroImage,* 33: 599-608 (2006); [3] Skudlarski et al., *NeuroImage,* 9: 311-329 (1999); [4] Maitra et al., *Magn Reson Med,* 48: 62-70 (2002).