A Novel Method for Robust Automated Thresholding in Pre-surgical fMRI using a Single Functional Run.

Tynan Stevens[1,2], David Clarke[3,4], Ryan D'Arcy[5,6], Gerhard Stroink[1], and Steven Beyea[2,7]

[1]Physics, Dalhousie University, Halifax, NS, Canada, [2]Neuroimaging Research Lab, BIOTIC, Halifax, NS, Canada, [3]Surgery, Dalhousie University, Halifax, NS, Canada, [4]Neurosurgery, QEII Health Sciences Centre, Halifax, NS, Canada, [5]Applied Science, Simon Frasier University, Burnaby, BC, Canada, [6]Surrey Memorial Hospital, Surrey, BC, Canada, [7]Radiology, Dalhousie University, Halifax, NS, Canada

Audience: Functional MRI users interested in automated analysis pipelines, specifically designed to produce robust thresholded activation maps at the individual level.

Background: We recently demonstrated a method for automated thresholding of test-retest fMRI images using ROC-reliability analysis (ROC-r)[1]. This method was shown to produce robust fMRI activation maps at the individual level. The main drawback of the test-retest ROC-r method is the requirement of repeated imaging, which a) introduces experimental constraints (i.e. limits the applicability of the method) and b) doubled scan times, and associated costs.

For this reason we have developed a new method of ROC-r threshold optimization that requires only a single fMRI run. This method uses split-halves subdivision of the fMRI time-series to produce an arbitrary number ($N_{iteration}$) of pseudo-test-retest activation map pairs and $2N_{iteration}$ estimates of the optimal ROC-r thresholds. The mean of these thresholds is applied to the mean of the pseudo-test-retest activation maps, in what is referred to as '1ROC-r' (as opposed to test-retest ROC-r, or 2ROC-r). The chief competing method for automated thresholding of single-run fMRI datasets is adaptive thresholding (AT)[2]. In the AT method the activation map histogram is modelled as a mixture of gamma distributions for activation and deactivation, and a gaussian distribution for noise. The threshold is taken as the crossing point of the activation and noise distributions. This method was validated by demonstrating high test-retest overlap.

Purpose: In this work, we compared the thresholds produced by the novel 1ROC-r method to both 2ROC-r, and AT. Furthermore, we assess the test-retest overlap produced by each thresholding strategy. We show that 1ROC-r produces better results than 2ROC-r, without the need for test-retest imaging. The ROC-r framework incorporates spatial information that is not utilized in the AT method, an advantage we believe is particularly relevant for noisy datasets (e.g. unexpected histogram content or low CNR) such as those encountered in pre-surgical language mapping. We will show that 1ROC-r provides a powerful and flexible means of automated thresholding in challenging fMRI contexts.

Methods: Sixteen individuals with cortical tumors volunteered to participate in this study (6 males, 10 females, 40 +/- 14 years of age). All volunteers were scanned with a 4 T scanner (Varian INOVA) while performing a variety of tasks. Tasks were chosen based on tumor location, including finger tapping (left or right hand, n=11), tongue movement (n=3), overt object naming (n=12), and sentence reading (n=8). For each task within session test-retest imaging was performed. Functional images were collected with a spiral out sequence (TR = 2 s, TE=15 ms, α=90°, 64 x 64 matrix, 22-25 slices, 3.75 x 3.75 x 4.5 mm voxels, 0.5 mm gap). A structural image was also collected using an MP-FLASH sequence (TI=500 ms, TR=10 ms, TE=5 ms, α=11°, 256 x 256 matrix, 190 slices, 0.94 x 0.94 x 1 mm voxels).

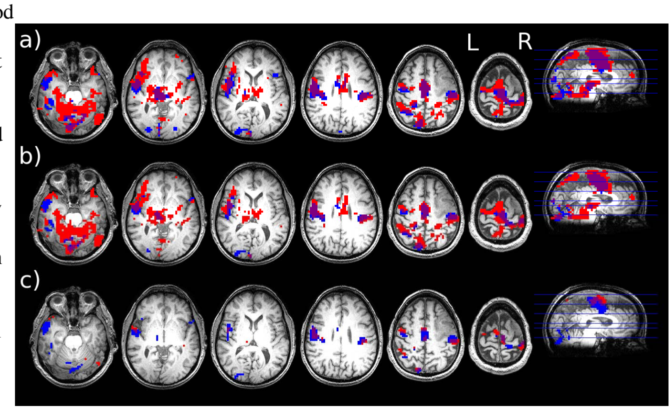

Figure 1: Test (red) and retest (blue) for a representative dataset (overt object naming) thresholded using the three automated methods: a) 1ROC-r; b) AT; c) 2ROC-r. The 2ROC-r thresholds were higher than AT or 1ROC-r thresholds, and 1ROC-r has the highest test-retest overlap (purple).

Standardized pre-processing steps included rigid body motion correction (data are excluded if > 2mm, n=3), registration to anatomical space, and brain extraction. For all other pre-processing steps, individualized pipelines were employed following the NPAIRS-GLM framework[3]. Each fMRI run was therefore subdivided into random split-halves 8 times, producing 16 activation maps per fMRI run, per pipeline. Pipelines tested included all combinations of: a) 0, 3, 6, or 9 mm smoothing, b) with or without auto-correlation correction, c) with or without motion parameter regression, and d) FSL or AFNI software (32 total possible combinations). Default parameters (6mm smoothing, no auto-correlation correction, no motion regression, AFNI) were used unless another pipeline produced significantly higher NPAIRS scores. The 16 activation maps from the best pipeline and their mean were retained for subsequent analysis.

The 8 pseudo-test-retest image pairs were used to calculate the 1ROC-r thresholds (i.e. the mean of the 16 pseudo-test-retest ROC-r thresholds), which was applied to the mean of the 16 images. This was then repeated for the other image of the test-retest pair. The 2ROC-r thresholds were computed directly from the mean test-retest images, and the AT thresholds were computed from the individual mean images independently. For all final thresholded test-retest maps, the Rombouts overlap coefficient was calculated and compared[4].

Results: The resulting thresholds were: 4.8+/-0.1 for 1ROC-r, 7.1+/-0.3 for AT, and 4.8+/-0.5 for 2ROC-r. The 2ROC-r (n=9) and AT (n=4) methods produced no reliable activation in more datasets than 1ROC-r (n=3). Two of these datasets produced no reliable activation by any method tested, and were excluded. An example of the thresholded activation maps for each method are shown in figure 1. The test-retest overlap for 1ROC-r (0.34+/-0.03) and AT (0.33+/-0.04) were significantly higher than 2ROC-r (0.21+/-0.04). For the motor datasets alone (n=13), test-retest overlap was 0.44+/-0.01 for AT and 0.39+/-0.02 for 1ROC-r, both of which outperform 2ROC-r (0.28+/-0.02). However, for language maps (n=16), the test-retest overlap of 1ROC-r (0.30+/-0.04) exceeds either 2ROC-r or AT (0.16+/-0.05 and 0.24+/-0.04). These results are summarized in figure 2.

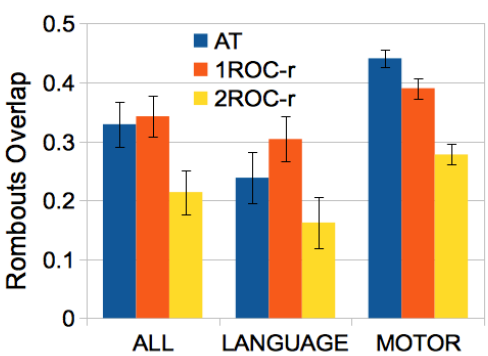

Figure 2: Test-retest overlap by threshold method and fMRI task. The 1ROC-r method had the highest overlap overall and was most consistent across tasks.

Discussion: 1ROC-r and AT thresholding produce identical threshold levels, and were not significantly different in terms of average test-retest overlap. However, 1ROC-r provides more consistent test-retest overlap across tasks, outperforming AT for language mapping, while taking a slight hit on motor tasks. This consistency of 1ROC-r is desirable for clinical applications like pre-surgical mapping, and enhances test-retest reliability from a single fMRI run.

The strengths of the 1ROC-r method include a) the use of spatial information for improved detection in noisy datasets (e.g. overt language tasks with patient populations), and b) availability of many pseudo-test-retest pairs for improved ROC estimation. Although 1ROC-r has substantial computational requirements, this is not seen as a formidable obstacle given trends in computing power. That 1ROC-r and AT arrive at similar threshold estimates and test-retest reliability is worth noting, and suggests that the combination of histogram fitting and reliability analysis may be complementary for automated threshold selection.

Conclusion: We have successfully demonstrated 1ROC-r for optimization of fMRI thresholds using only a single run of fMRI data. The 1ROC-r method produces more robust automated thresholds, and is particularly successful in difficult applications like presurgical language mapping.

References: [1]Stevens et al 2013, J. Neurosci. Meth. **219**, 312; [2]Gorgolewski et al 2012, Front. Hum. Neurosci. **6**, 245; [3]Kjems et al 2002, NeuroImage **15**, 772; [4]Rombouts et al 1997, Am. J. Neuroradiol. **18**, 1317.