Reproducibility of fMRI Brain Maps for Presurgical Planning: Final Report

J. Mathis & E. DeYoe

Subaward No: HHSN268201000050C (9a)
PI: E. DeYoe
Reporting Period: Final
Project end date (revised): 9/30/12

Contact:

E. DeYoe
Medical College of Wisconsin
Dept. Radiology
8701 Watertown Plank Rd.
Milwaukee, WI 53226
414-456-4920
deyoe@mcw.edu

Project Deliverable Milestones (original)**:**

1) Month 0: Sign award agreement. Complete IRB approval.
2) Month 3
   - Select, test and install standardized computational sequences including the AMPLE algorithm coordinated with Voyvodic project.
   - Data compilation/Preprocessing: Data for 8 subjects organized in local imaging data base with table entries to facilitate scripted data queries and analysis. Registration of MRI scans with standard atlas brain, generation of brain activation maps, and calculation of image-based quality assurance QA metrics (image motion, signal drift, signal spikes, task activation).
3) Month 6
   – Calculate reproducibility metrics for all repeat scans. This includes: calculation of AMPLE normalization parameters (both atlas-based and automated activation cluster-based ROI's), calculation of AMPLE temporal stability metrics, calculation of reproducibility metrics across repeat scans (variability in laterality index, location of active clusters, and spatial extent of active clusters).
4) Month 12:
   – Develop and apply repeatability measurements for functional specificity for 8 datasets
   - Meta-analyses with Voyvodic subproject.
   - Prepare report including quantitative metrics of scan-rescan reproducibility plus description of methodology and QA procedures used to achieve the observed levels of reproducibility. Coordinate report with Dr. Voyvodic's report. Using the report results, draft a Reproducibility Claim for the QIBA fMRI Profile.
   - Prepare data set for inclusion in a QIBA data archive.

At the time of this writing the Deliverable Milestones for this project have been completed with the following modifications (using the above numbering):

3) The *laterality index* was determined not to be relevant for the vision fMRI datasets analyzed in this subproject since the data do not exhibit lateralized differences.

4) The *meta-analysis* in which the data from this subproject and the Voyvodic subproject were to be pooled may not be practical at least as originally envisioned. An important finding of this subproject is that reproducibility metrics may yield significantly different results depending on the type of fMRI data being analyzed. This is important since a goal of the two subprojects was to use methods that were as identical as possible. Despite the matched methodology, the results for the analysis of fMRI vision mapping data appears to yield results that differ in important respects from those obtained for hand-motor and language mapping. At this juncture, the primary difference appears to be in the effect of changes in statistic threshold on reproducibility. For vision mapping data, reproducibility is slowly degraded as the statistical criterion becomes more stringent and effectively reduces the number of voxels (amount of brain tissue) that is considered active. This may not be the case for simple motor mapping of a single extremity (eg hand), perhaps because the focus of activity in the latter case consists of a smaller, discrete focus of activity. In contrast, visual field mapping activates a broad swath of occipital cortex extending across multiple, functionally distinct but contiguous visual areas. In such case, the broad expanse of activation that is unbroken at low statistical threshold can break up into multiple, separate foci at higher thresholds thereby reducing the apparent reproducibility of measures such as the center-of-mass. In sum, the current results suggest that there are factors that may differ between datasets from different modalities or behavioral tasks that may prevent direct pooling of results without regard to those factors. Also, the original proposal to submit the data used in this project to a QIBA data archive awaits further

development of that archive and specification of the format in which data are to be submitted. Once this has been clarified, we will pursue this milestone.

Summary of Results:

The following provides a summary of the primary results of this project that bear directly on the fMRI profile claims being developed by the QIBA fMRI technical committee.

ROI definition – visual cortex:

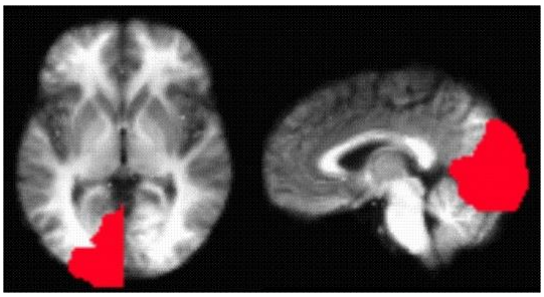

Figure 1

A region-of-interest (ROI) for occipital visual cortex in each hemisphere was established using a combination of 2 components (termed “cuneus” and “lingual gyrus”) defined for the Montreal Neurological Institute’s ICBM 152 brain atlas. Starting with this ROI, we then performed an overlap analysis to determine a new ROI that included all voxels that were classified as active at each of 5 T-statistic thresholds (3,4,6,8,10) and three AMPLE thresholds (40%,60%,80%). We then regularized the boundary of this ROI by a 2-step boundary decimation followed by a 3-step boundary expansion. As illustrated in Figure 1 for the left hemisphere, this yielded a final occipital ROI for each hemisphere that ensures that the ROI does not artificially limit the extent of activation even at the most liberal statistical threshold. All subsequent analyses were performed on the activation contained within these final ROIs.

Data acquisition and processing overview:

Visual cortex, t2*-weighted, fMRI activation was obtained from 8 healthy volunteers at 3 Tesla using a conventional temporal phase mapping paradigm with both an expanding, checkered ring and a rotating checkered quarter wedge[1, 2]. DICOM fMRI and T1-weighted anatomical images from each MRI scan were not motion corrected in order to match the procedure used by Dr. Voyvodic[3, 4]. fMRI timecourse data were spatially smoothed in 3 dimensions with a 6 mm full-width-half-maximum spherical kernel before alignment to the ICBM 152 atlas using AFNI’s 3dAlineate function. AFNI’s 3dFim+ and 3dCalc algorithms were used to compute a T statistic for each voxel. AMPLE normalization was then performed using the method described by Dr. Voyvodic with the modification that the ROI described above was used to identify the maximum T values in each hemisphere which were then used for normalization in each hemisphere separately.

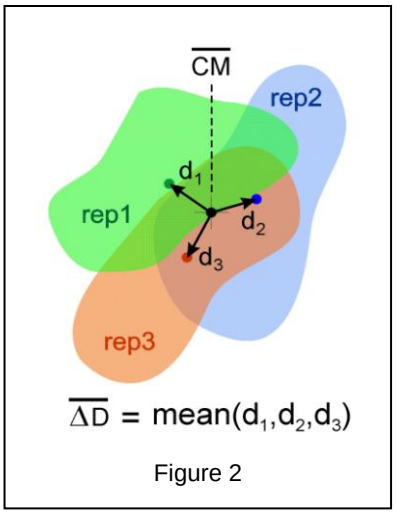

Figure 2

Reproducibility metric – weighted center of mass (wCM):

Figure 2 and the following equations describe the computation of the weighted center-of-mass metric. Dr. Voyvodic has argued that the weighted center-of-mass provides a more reproducible metric than the un-weighted center-of-mass because it tends to emphasize those brain voxels with the strongest signals.

Within day – having N repeated measurements [x_i y_i z_i T_i], where T = T or AMPLE value

(1) $X = \frac{\sum_i^N x_i * T_i}{\sum_i^N T_i}$ ... *Center of Mass cords, same for Y, Z*

(2) $\overline{CM} = X\ Y\ Z$ *Mean coordinates within/across day*

(3) $[\Delta x_i\ \Delta y_i\ \Delta z_i] = x_i\ y_i\ z_i - \overline{X}\ \overline{Y}\ \overline{Z}$ *Difference from within/across day mean*

(4) $\Delta d_i = \overline{\Delta x_i^2 + \Delta y_i^2 + \Delta z_i^2}$ *Distance from within/across day mean*

(5) $\overline{\Delta D} = \sum_i^N \Delta d_i / N$ *Mean distance deviation*

Across day - As above except N is replaced by M = # days
(There will be an across day measure for each within day repetition and for each subject)

Figure 2 provides a visual representation of the preceding method for measurement of the reproducibility metric for the center-of-mass across repeated observations within the same MRI scan session (Within-Day) and across multiple scans sessions on different days (Across-Day). Table 1 illustrates the organization of the data. Each observation, [xyzT] represents the average distance deviation, $\overline{\Delta D}$, from either the within-day (orange) or across-day (green) mean weighted center-of-mass, $\overline{CM}$ of the suprathreshold activation within the analysis ROI.

	Rep 1	Rep2	Rep3	Rep4		Mean x y z
Day 1	[x y z T]	[x y z T]	[x y z T]		=>	Within day
Day 2	[x y z T]	[x y z T]	[x y z T]		=>	Within day
Day 3	[x y z T]	[x y z T]	[x y z T]	[x y z T]	=>	Within day

	Rep 1	Rep2	Rep3	Rep4
Day 1	[x y z T]	[x y z T]	[x y z T]	
Day 2	[x y z T]	[x y z T]	[x y z T]	
Day 3	[x y z T]	[x y z T]	[x y z T]	[x y z T]
	V	V	V	
	Across day	Across day	Across day	
	Mean x y z			

Table 1

Figure 3 and associated Table 2 show the reproducibility metric for the center-of-mass, $\overline{\Delta D}$, for a range of statistical threshold settings for conventional T-valued data and AMPLE normalized data for both within-day and across-day repeated measurements. Since the statistical thresholds for T and AMPLE data are not directly comparable, we used the method illustrated in Figure 4 to establish threshold settings that yielded approximately equal numbers of voxels classified as active for the two measures. The relative locations of the T and AMPLE curves in Figure 3 reflect this approximate equivalence. For both measures, the reproducibility of the center-of-mass location for activation in visual cortex is good (< 5 mm deviation) for low to moderate thresholds but degrades as the statistical criterion becomes more stringent. Note that at the highest thresholds, the total number of voxels

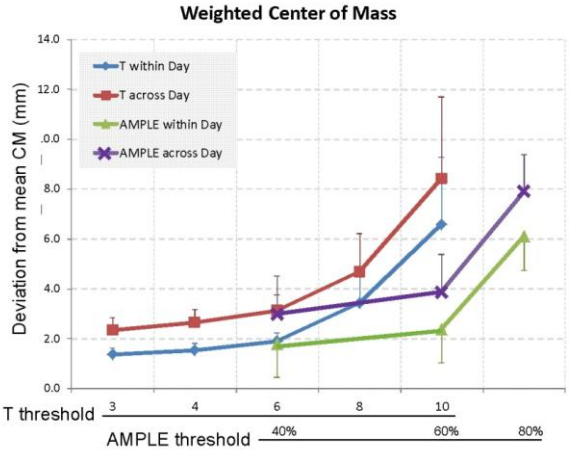

Figure 3

(Fig. 4) in the activation pattern becomes relatively sparse. Consequently, variation in the locations of a small number of voxels can have a large impact on the location of the center-of-mass. Also note that for both T and AMPLE data, reproducibility is best for repeated measurements obtained within-day without moving the subject between repetitions. Even so, the across-day repeatability is only about 1 mm worse when the subject's head is in a different position from day-to-day. This suggests that the computational alignment of data to the ICBM atlas using AFNI's 3dAlineate has a relatively small effect on reproducibility especially since there are likely to be other factors that contribute to the slightly poorer reproducibility across days compared to within-day.

All Subjects		T		AMPLE	
		W-D	A-D	W-D	A-D
Threshold	3	1.38	2.35		
	4	1.55	2.64		
	6	1.90	3.13	1.79	3.02
	8	3.43	4.70		
	10	6.57	8.42	2.37	3.88
	12			6.09	7.89

Table 2

In sum, at T thresholds of 8.0 or less and AMPLE thresholds of 60% or less, the weighted center-of-mass is reproducible to within 5 mm across days and to within 4 mm within day. Also, at equivalent threshold settings, AMPLE normalization provides equivalent or better reproducibility. An AMPLE threshold of 60% which has been recommended by Dr. Voyvodic, may provide an optimal combination of discreteness plus reproducibility within the limits specified above.

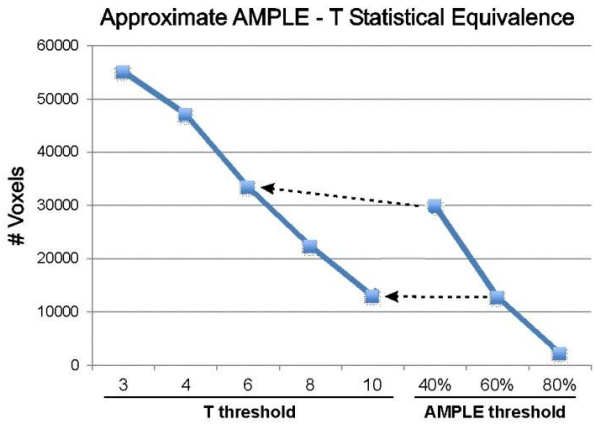

Figure 4

***Reproducibility metric – # Active Voxels*:**

Figure 5 and associated Table 3 illustrate results for the reproducibility metric of the number of active voxels contained within the fMRI activation pattern. Here the metric consists of the percentage deviation in the mean number of voxels. Since the absolute number of voxels varies considerably with the threshold, expressing the standard deviation as a percentage of the mean provides a measure that can be meaningfully compared across the various threshold settings. Figure 5 shows that the % deviation increases steadily as the threshold becomes more stringent for both T and AMPLE data. Again, the AMPLE normalization provides improved reproducibility with an AMPLE threshold of 60% providing a potentially optimal compromise between discreteness and reproducibility. As for the center-of-mass metric, the repeatability across days is worse (larger deviation) than for with-session measurements.

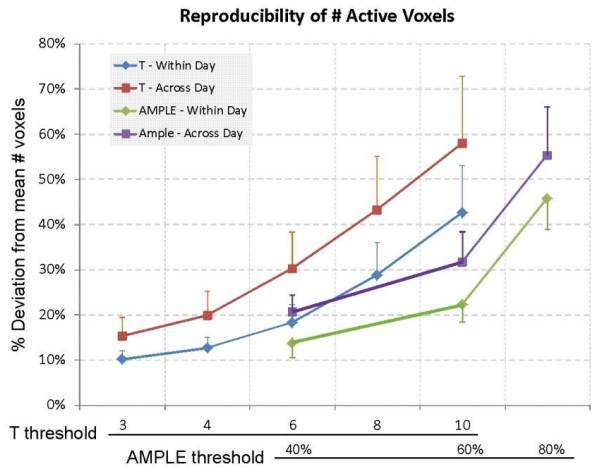

Figure 5

In sum, for T thresholds of 6.0 or lower and

AMPLE thresholds of 60% or less, the mean number of voxels within the fMRI focus is reproducible to within 30% across days and to within 20% for measurements repeated within the same scan session.

All Subjects		T		AMPLE	
		W-D	A-D	W-D	A-D
Threshold	3	10%	15%		
	4	13%	20%		
	6	18%	30%	14%	21%
	8	29%	43%		
	10	43%	58%	22%	32%
	12			46%	55%

Table 3

***Reproducibility metric – % Overlap*:**

Though conceptually simple, the percentage overlap measure has certain interpretational subtleties. In Figure 1 the percentage of voxels from pattern "rep1" that overlap with "rep2" appears to be approximately 33% whereas more than 50% of "rep3" overlaps with "rep2". Note, however, that the measure can be "directional" such that if one pattern is significantly smaller than another, their % overlaps can be quite different. In other words if pattern A is, say, twice the size of pattern B, only about 25% of A might overlap B but then 50% of B would overlap A. Our analysis provided both A/B and B/A overlap measures but here we report just the larger of the overlap measures since this is arguably the more relevant of the two measures. For instance, in the preceding example, if B overlaps A 100% (in which case A only overlaps B by 50%), the interpretation would be that the two patterns are largely concident whereas the lower measure would incorrectly imply only modest conicidence.

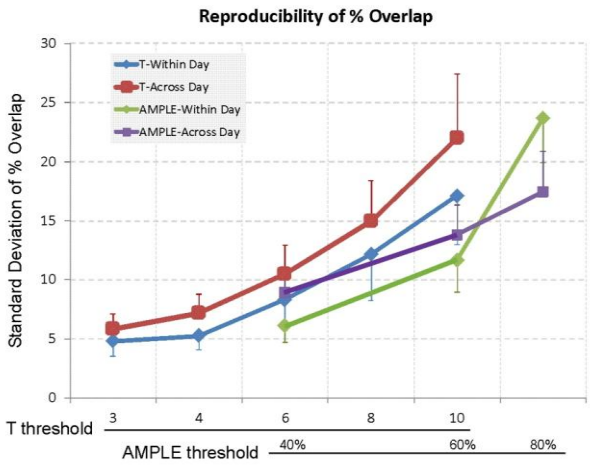

Figure 5

Figure 5 and associated Table 4 summarize the results for the reproducibility of the % overlap measure. As the statistical threshold becomes more stringent, the reproducibility degrades steadily for both conventional T values and AMPLE normalized values with the latter maintaining better reproducibility at matched statistical thresholds. Within-day measures were also slightly better than across-day measures by about 2-5%. In summary, at T thresholds of 8 or less and AMPLE thresholds of 60% or less, the % overlap of an fMRI focus is repeatable to within 15% across days and to within about 12% for repeated measures in the same session.

All Subjects		T		AMPLE	
		W-D	A-D	W-D	A-D
Threshold	3	5	6		
	4	5	7		
	6	8	10	6	9
	8	12	15		
	10	17	22	12	14
	12			24	17

Table 4

Discussion:

Overall, the results of this study indicate that visual cortex fMRI activation foci are, on the whole, quite reproducible with respect to the location of the wieghted center-of-mass, the number of voxels within the focus and the spatial overlap of repeated observations. AMPLE normalization can improve repeatability relative to conventional T-valued data at

comparable statistical thresholds. However, anecdotal observations with other visual mapping data not presented here suggest that AMPLE normalization may at times suppress weak activation in portions of a broad pattern of activity (eg extrastriate cortex) that might be important for clinical interpretation in the context of presurgical planning.

A manuscript for journal publication is in preparation that will describe additional details and results from this study that go beyond the items described above which most directly impact the claims being considered for inclusion in the QIBA fMRI profile.

References:

1. DeYoe, E.A., et al., Mapping striate and extrastriate visual areas in human cerebral cortex. Proceedings of the National Academy of Sciences - USA, 1996. 93(6): p. 2382-2386.
2. DeYoe, E.A., et al., *FMRI of Human Visual Pathways*, in *Functional Neuroradiology: Principles and Clinical Applications*, S. Faro and F.B. Mohamed, Editors. 2011, Springer: New York. p. 485-511.
3. Voyvodic, J.T., *Activation mapping as a percentage of local excitation: fMRI stability within scans, between scans and across field strengths.* Magn Reson Imaging, 2006. **24**(9): p. 1249-61.
4. Voyvodic, J.T., J.R. Petrella, and A.H. Friedman, *fMRI activation mapping as a percentage of local excitation: consistent presurgical motor maps without threshold adjustment.* J Magn Reson Imaging, 2009. **29**(4): p. 751-9.