QIBA Round-1 Project 10a Final Report

Sept 2012

fMRI Project: **Quantitative measures of fMRI reproducibility for pre-surgical planning**
Voyvodic Sub-project: **Development of reproducibility metrics**

This sub-project started 7/15/11 and ended in July 2012. We have successfully completed our project's **Aim** and our analysis has tested its 3 **Hypotheses**, which were:

Aim: To develop metrics for quantifying reproducibility within and across fMRI scans and to apply those metrics to assess the reproducibility of results obtained using alternative fMRI analysis methods on existing data sets. Initial analysis will focus on repeated scan data using the AMPLE normalization algorithm (Voyvodic, 2006) to improve and enable quantification of reproducibility. We will also generate quantitative data quality metrics for each individual data set (e.g. SNR, motion, AMPLE temporal stability). Subsequent analyses will generate QA metrics and apply AMPLE normalization to patient data sets to assess intra-session and inter-subject reproducibility.

Hypothesis 1: AMPLE normalized maps of language, motor, and primary sensory brain areas will be quantitatively reproducible -- the study will determine empirical measures of that reproducibility

Hypothesis 2: AMPLE temporal stability metrics can be used to predict scan/rescan reproducibility

Hypothesis 3: for single session patient data, AMPLE temporal stability metrics will be correlated with other quantitative QA measures and with subjective ratings of confidence in functional localization.

During this study we accomplished the following in addressing our **Aim**:

1) All data sets in the 3 data collections were organized in a relational imaging database containing clinical and behavioral meta-data for each scan session, with direct links to the associated imaging data. The 3 data collections were: 1) healthy controls and 2) patients performing a clinical fMRI protocol, and 3) FBIRN subjects performing a multi-site comparison protocol. Our local image processing software (fScan) can query the database to find image data sets with particular properties and then use the file pointers from the database to access the imaging data directly. We have improved these features and created automated analysis scripts to perform standard preprocessing on all the reproducibility data sets, and to extract activation metrics for each scan.

2) We generated AMPLE normalized activation maps for all subjects who underwent multiple scan sessions in our local data collections and the FBIRN Phase 1 data. For the FBIRN data we found that AMPLE normalization improved reproducibility between scans acquired at different sites, but there remained considerable inter-session variability, probably due to the large amount of heterogeneity of scanner types, pulse sequences, and procedures used at different sites. We focused our quantitative analysis therefore on the language mapping data sets in our Duke collections. We calculated activation metrics (cluster location, spatial extent, and hemispheric dominance) for all language maps. From these we calculated reproducibility metrics (shift in peak location, percent activation overlap, and percent change in laterality index) by pair-wise comparisons of similar language scans collected for the same subject in different sessions to test Hypothesis 1.

3) We developed automated processing scripts to analyze AMPLE temporal stability of functional data sets. These automated scripts were applied to all language reproducibility data sets to test the prediction that AMPLE stability metrics will be strongly correlated with fMRI scan quality and reproducibility (Hypotheses 2 & 3).

4) Both the Deyoe lab in Wisconsin and the Pillai lab in Maryland installed our fScan software and used it to perform AMPLE normalization analyses on their own data sets. They have also implemented the AMPLE algorithm into their own software.

Results

This project has produced the following results:

1) AMPLE normalization did greatly increase quantitative reproducibility of fMRI language mapping, as illustrated in Figure 1, which compares standard t-maps and AMPLE maps for a single subject scanned 6 times. For all subjects tested, the spatial extent of activation for either motor or language tasks was highly variable in standard statistical t-maps, and much more reproducible after AMPLE normalization during image post-processing. These results support **Hypothesis 1**.

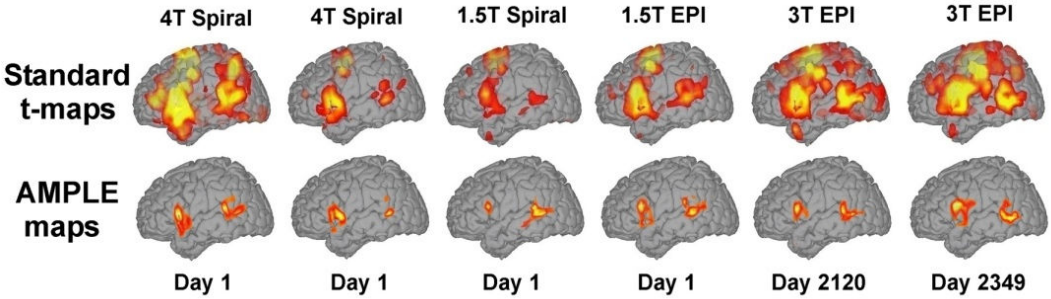

Figure 1 Reproducibility of AMPLE fMRI mapping. Comparison of fixed-threshold t-maps versus AMPLE maps for frontal and temporoparietal language activations for a single subject scanned 6 times under different conditions. Days listed below are relative to the date of the first scan.

2) Quantitative measures of reproducibility showed that the brain location for the peak of activation in AMPLE-normalized maps was highly reproducible across scan sessions. We found that peak location of language areas varied by less than 10mm in almost all pair-wise comparisons, and by less than 5mm when both scans were acquired using the same scanner model and the same pulse sequence (**Figure 2A**). Reproducibility of the spatial extent of activation was not correlated with acquisition procedures, but was positively correlated with the strength of the task activation signal itself (**Figure 2B**). Reproducibility of hemispheric laterality index in AMPLE normalized maps was very good, with over 90% agreement between scans in frontal and temporoparietal language areas (not shown). These results support **Hypothesis 3**.

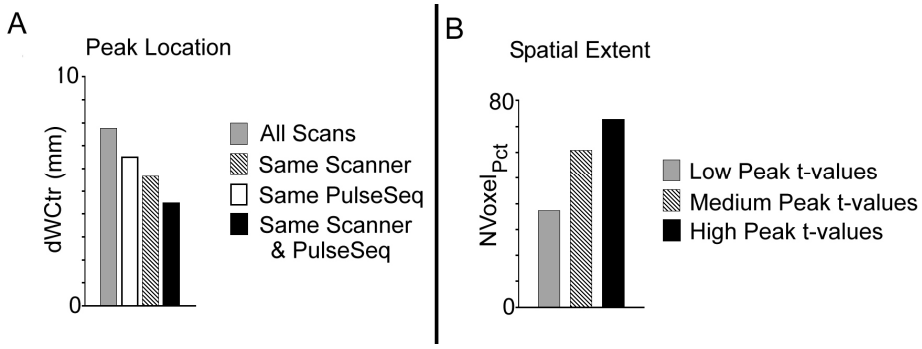

Figure 2 Reproducibility of activated area peak location and spatial extent. **A)** Peak location reproducibility was correlated with similarity of acquisition methods, being best for scans acquired on the same scanner using the same pulse sequence. **B)** Reproducibility of the spatial extent of activation was correlated with the amplitude of the activation signal itself, being best for strong task activations.

The results above have been recently published in JMRI (Voyvodic, 2012).

3) Our analysis of the temporal consistency of fMRI language mapping found that the spatial extent of activation did tend to stabilize as a function of scan time, consistent with our earlier studies (Voyvodic, 2006; Voyvodic et al., 2009). In pair-wise comparisons of repeat scans, however, we did not find that the reproducibility of either the peak location, nor the spatial extent of language activations was correlated with how quickly the AMPLE plots reached stable plateau levels. This finding appears to contradict our **Hypothesis 2**. Further study will be needed to examine this issue.

4) Analysis of these reproducibility data lead to some preliminary results that went beyond the scope of our original goals. In particular, whereas our fMRI Profile has this far focused on quantifying the location of brain active areas, we have found that the amplitude of BOLD activation signals can also be quite reproducible across scan sessions. Recent analyses emerging from discussions with Dr. DeYoe have shown that differences in BOLD signal amplitude in response to different levels of task load are consistent across trials within a scan, and the ratio of those responses is similar when compared across different scan sessions. This raises the intriguing possibility that quantifying responses to different task loads within each session may provide a way to calibrate quantitative changes in the amplitude of fMRI responses across different sessions. Having such a calibration mechanism could in principle be very useful as a quality assessment tool for qualifying fMRI scans. It would also help move fMRI beyond a method for quantifying the location of brain active areas, to become a method for quantifying the amplitude of brain activity. We plan to continue exploring these preliminary observations in further studies.

5) As a direct offshoot of this QIBA-supported project I have prepared an NIH U01 grant application entitled “Quantitative imaging of brain function and connectivity”, which will be submitted in Oct 2012 in response to the Quantitative Imaging Network initiative of the NCI. This study, if funded, will address all of the issues raised in this QIBA project in much greater depth. Preliminary results provided by the QIBA project were instrumental in preparing this larger study.

6) The data sets used in this project could be made available for sharing via QIBA’s Open Image Archive Initiative. The logistics of actually submitting imaging data and associated metadata have not yet been worked out, however, and as a result we have not yet attempted to prepare these data sets for export. As these retrospective data sets were not acquired specifically to address quantitative reproducibility issues, it is our expectation that uploading them would be of limited use. New data sets should be acquired using a QIBA supported fMRI protocol, incorporating our recent findings, which could then provide publicly available support for the claims in our Profile.

QIBA-supported publication:

J.T. Voyvodic (2012) “Reproducibility of single-subject fMRI language mapping with AMPLE normalization”, J. Mag. Res. Imaging, 36:569-80.

Related literature cited:

J.T. Voyvodic (2006). Activation mapping as percentage of local excitation (AMPLE): fMRI stability within scans, between scans, and across field strengths, Magnetic Resonance Imaging, 24:1249-1261.

J.T. Voyvodic, J.R. Petrella, and A.H. Friedman (2009) “fMRI activation mapping as percentage of local excitation: Consistent presurgical motor maps without threshold adjustment”, J. Mag. Res. Imaging 29:751-759.