Test-retest Reliability of functional MRI using Smart Phantom: Analysis II

Q. Zhao[1], T. Conway[2,3], K. White[2,3], K. McGregor[2,3], A. Ford[2,3], and B. Crosson[2,3]

[1]University of Georgia, Athens, GA., United States, [2]University of Florida, Gainesville, FL., United States, [3]VA Medical Center, Gainesville, FL.

Introduction

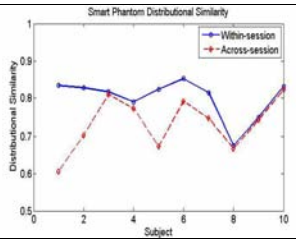
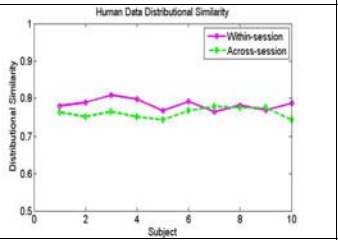
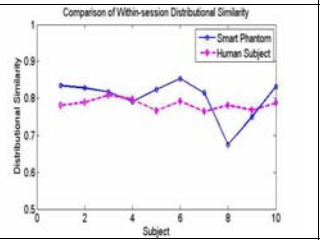
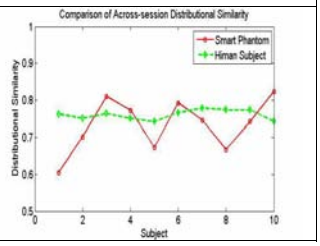
fMRI uses blood oxygenation level dependent (BOLD) contrast to construct task-related activation maps for the human brain. However, repeatable estimates are difficult to obtain because of many sources of variability, such as human subjects' anatomical differences, physiological variations, irreproducible motion artifacts, etc. Smart Phantom [1,2], a device that eliminates those sources of variability from reliability tests by generating reproducible simulated BOLD activation signals, has been used for a test-retest reliability study with initial results reported [3]. In this presentation, further detailed results of test-retest reliability will be reported.

Data Acquisition & Analysis Methods

Two sets of data, Smart Phantom's and human subjects', were acquired with a Siemens 3T Allegra head-only scanner. Each time series contained 125 images and 10 events, taking 213 sec to acquire. For Smart Phantom data, at each event a triangle waveform was generated in the Smart Phantom coils, which roughly simulated the temporal frequency content of a BOLD hemodynamic response (HDR) having an amplitude approximating a 3% signal change. There were 12 images per event and 10 events per run, which produced simulated HDRs similar to the actual HDRs of an adult with post-stroke aphasia. The Smart Phantom scan parameters were: TR=1700ms, TE=25ms, FA=70, FOV=240mm matrix 64x64, 25 slices with 5mm thickness, no gap between slices (acquisition voxels were approximately 3.8 x 3.8 x 5 mm). During each human subject scanning session, the Smart Phantom was scanned twice. One pre-subject Smart Phantom run of 125 images occurred before the human fMRI runs and one post-subject Smart Phantom run of 125 images occurred after the human fMRI runs. Human data were obtained from two populations of adults, eight controls and five patients with chronic aphasia due to stroke, scanned in thirty-nine sessions (thirteen subjects, three sessions per subject) over a period of about 5 months. The scanning sessions occurred at fixed intervals: initial scan, 1-week post initial scan, and 12-weeks post initial scan. The human scan parameters were identical to the Smart Phantom parameters, except 32 slices with 5mm thickness and no gap between slices were acquired to cover the whole brain. In each session, 5 separate runs were performed for each of two language-related tasks, reading aloud visually presented pseudo-words or repeating aurally presented pseudo-words. Each participant was scanned repeatedly at the same time of day and day of week. During each scanning session participants performed the same tasks, both within and across sessions. Analyses of simulated and actual HDRs were performed with AFNI software, including 3Dvolume registration, deconvolution, cross-correlation, and removal of speech-related motion artifacts from human data via selective de-trending. Correlation maps were calculated by cross-correlating each brain voxel time series with an ideal signal, for Smart Phantom a triangle wave convolved with the event time course, and for human data an idealized HDR (provided by AFNI) convolved with the event time course. For both the Smart Phantom and human data, the Integrated Squared Error (ISE) [4] was used to calculate the distributional similarity of two activation maps as the reliability index. Distributional similarity has the advantage of being relatively insensitive to in-plane rotations and translations of the images.

Results and Discussions

For Smart Phantom scans, within-session distributional similarity was analyzed comparing the pre-subject run to the post-subject run in the same session(s). Across-session distributional similarity was analyzed using all possible combinations of a pre-subject run vs. a post-subject run, from any two of the three sessions. For human scans, within-session distributional similarity compared data acquired during run 1 and run 5 of each task per session. Across-session distributional similarity compared data from the same runs across any two of the three sessions. Figure 1 (left) compares Smart Phantom's within-session and across-session distributional similarities, while Figure 1 (right) compares human subjects' within-session and across-session distributional similarities. On average, and not surprisingly, within-session distributional similarity is better than across-session for both Smart Phantom and human data. Comparing within-session distributional similarity for Smart Phantom to human data, shown in Figure 2(left), shows higher similarity for Smart Phantom in all but three cases. While this is the expected direction, the findings are less consistent than we had expected to find. A contrasting picture emerges comparing across-session distributional similarity for Smart Phantom to human data, in Figure 2(right).This shows higher similarity for Smart Phantom in only four cases out of ten, that is, the data are more often in the unexpected direction. Clearly, Smart Phantom distributional similarity indices present higher variation than do the corresponding human indices. One possible cause for greater Smart Phantom variation is inconsistent positioning of the Smart Phantom in the head coil and/or the main magnetic field, due to the lack of a calibrated phantom holder. A second, related possible cause is that the Smart Phantom is spherical and without any internal landmarks, which makes it difficult to co-register the 3 sessions' scans. It is much easier to co-register a human subject's brain images using anatomical landmarks, and image registration partially mitigates inconsistent head positioning.

Figure 1 (left)	Figure 1 (right)	Figure 2 (left)	Figure 2 (right)
			

In general, the Smart Phantom can generate consistent BOLD "simulated activation" without introducing any physiological noise. In this sense, reliability estimates using the Smart Phantom may provide a gauge of hardware reliability for capturing BOLD-like signals absent the sources of fMRI variability arising from human subjects. However, additional refinements are necessary before reliability analyses using Smart Phantom will be more consistent over time than human subjects' data. Future research will explore the impact of a calibrated phantom positioning device to improve within- and across-session reliability.

REFERENCE:

1. Zhao, Q, Duensing GR, Fitzsimmons J. Proc. ISMRM (2003), 1843.
2. Cheng, H., Zhao, Q. et al. Magnetic Resonance Imaging. 24(2006): 301-313
3. Zhao, Q., White, KW, Conway, T., McGregor, K., Crosson, B. Proc. ISMRM (2006): 2835

- This material is based upon work supported in part by the Office of Research and Development, Rehabilitation R & D Service, Department of Veterans Affairs. Smart Phantom data was acquired with support of Invivo Corporation.