Available online at www.sciencedirect.com
ScienceDirect
Magnetic Resonance Imaging xx (2006) xxx–xxx
MAGNETIC RESONANCE IMAGING

Activation mapping as a percentage of local excitation: fMRI stability within scans, between scans and across field strengths

James T. Voyvodic*

Brain Imaging and Analysis Center, Box 3918 Duke University Medical Center, Durham, NC 27710, USA
Department of Radiology, Duke University Medical Center, Durham, NC 27710, USA

Received 7 November 2005; accepted 28 April 2006

Abstract

Functional magnetic resonance imaging (fMRI) does not typically yield highly reproducible maps of brain activation. Maps can vary significantly even with constant scanning parameters and consistent task performance conditions (Liu et al., Magn. Reson. Med., 2004, 52:751–760). Reproducibility is even more of a problem when comparing fMRI signal magnitude and spatial extent of activation across scans involving different task performance levels, scan durations, pulse sequences or magnetic field strengths. In this report, the consistency of fMRI was reexamined by considering the relative spatial and temporal distribution of fMRI blood oxygen level dependent (BOLD) activation signals separately from the absolute magnitude of the activation signal in each brain area. Subjects repeatedly performed the same simple motor task but under a variety of imaging conditions, using both spiral and standard echo-planar pulse sequences and at 1.5- and 4.0-T magnetic field strengths. The results demonstrate that the absolute amplitude of BOLD statistical activation signals varied significantly across time and scanning conditions, but the relative spatial pattern of BOLD activation was highly reproducible across all conditions. Analysis of realistic simulated fMRI data sets indicates that stability of relative activation patterns could provide a useful tool for assessing the accuracy of fMRI maps.

Keywords: Brain; Visualization; Image processing; Reproducibility

1. Introduction

Functional magnetic resonance imaging (fMRI) measures blood oxygen level dependent (BOLD) changes in signal intensity that are correlated with changes in brain activity [1,2]. fMRI mapping has been proven to be very useful for qualitative mapping of brain function, but it has been less successful for quantitative mapping of brain activity (reviewed in Refs. [3,4]). This is in part because the inherently indirect nature of the BOLD signal makes it difficult to determine the exact source or spatial extent of brain activity. In addition, quantitative fMRI mapping has been hindered by the fact that active brain voxels are typically identified based on statistical significance values rather than direct BOLD signal levels. As a result, the number of active voxels changes if the statistical threshold level is adjusted or if more trials are averaged together. Apparent activations also change if the MR signal-to-noise ratio is changed by using different pulse sequence parameters or by scanning at different magnetic field strengths.

Even under constant scanning conditions, the amplitude and spatial extent of fMRI activations can also be affected by variations in subject's task performance and attention levels. Other contributors to variability include fluctuations in scanner signal stability, as well as unconscious physiological variables such as heart rate and respiration. Liu et al. [5] recently reported that even when behavioral performance was carefully controlled to be the same across multiple runs of a single subject performing a simple motor task under constant imaging conditions, the spatial extent of activated voxels varied significantly across repeated fMRI scans. Given such variability, most fMRI studies depend on signal averaging across multiple voxels, runs and/or subjects imaged under constant scanning conditions in order to obtain statistically reliable maps for comparing different task conditions or subject populations [4].

* Brain Imaging and Analysis Center, Box 3918 Duke University Medical Center, Durham, NC 27710, USA. Tel.: +1 919 668 2609; fax: +1 919 681 7033.
E-mail address: jim.voyvodic@duke.edu.

0730-725X/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.mri.2006.04.020

Increasingly, however, fMRI is being used in applications where simple signal averaging for comparing relative activation maps is not sufficient. For example, in multisite research studies, different subjects are likely to be scanned at different field strengths and on different models of MRI scanners. Identifying calibration metrics to cope with site-to-site differences in fMRI activation maps could significantly enhance the statistical power of such studies [6]. Comparing scans using calibration metrics is limited, however, to scans for which such metrics are known, and it does not address scan variability due to differences among scans on a single scanner.

Variability in fMRI scans is also a serious problem for imaging applications that attempt to map brain activation within individual subjects, for whom signal averaging across many scan runs may not be practical. In clinical fMRI for neurosurgical planning, for example, the goal is to make reliable diagnostic maps of areas of essential brain function in each individual subject within a clinically realistic scanning session. Many studies have demonstrated the feasibility of diagnostic fMRI for identifying critical language or primary sensory and motor areas (reviewed in Ref. [7]) and have shown that fMRI active areas generally agree with functional localization based on intraoperative electrophysiological mapping. There are also many differences, however, and considerable variability across scans. Some of this variability is due to disease-related tissue pathology that interferes with the normal physiological mechanisms of the BOLD signal. Intersubject differences in task performance, cortical anatomy and neurovascular physiology are also likely to contribute to variability in fMRI results. Improved imaging methods and behavioral paradigms may help to reduce the amount of variability [6], but diagnostic fMRI also needs improved analysis metrics for describing the location and spatial extent of active brain areas that are relatively independent of scanning conditions or total imaging time.

This report addresses the issue of fMRI reproducibility by testing a spatially adaptive analysis approach aimed at identifying activation metrics that are relatively insensitive to interscan variability. The method is based on activation mapping as a percentage of local excitation (AMPLE). The rationale for the AMPLE approach is that the absolute statistical significance of BOLD responses is inherently dependent on task duration and scanner signal-to-noise sensitivity, whereas the relative pattern of signal amplitudes within localized clusters should depend primarily on the spatial distribution of the BOLD signal itself. Thus, for a given pattern of brain activity, the relative fluctuation of BOLD signal across the brain voxels involved in the task should remain fairly stable both within and across scans. The AMPLE approach, therefore, aims to create standard fMRI statistical activation maps and to create additional normalized voxel activation maps that characterize the relative spatial distribution of the BOLD statistical signal within each brain area. The results presented here show that, as predicted, when individual subjects performed a simple motor behavioral task under a variety of scanning conditions, absolute peak activation values were quite variable, whereas the spatial distribution of relative fMRI activation signals remained stable both within and across scans. Test scans and simulations demonstrate that the AMPLE method could provide a robust analysis tool to aid in evaluating the reproducibility and reliability of fMRI statistical activation maps.

2. Methods

2.1. Design

Human volunteer subjects performed a simple bilateral hand motor task a number of times while undergoing fMRI scanning under a variety of different scanning conditions. Magnetic field strength, pulse sequence and task performance rate were varied across scans in order to sample a range of fMRI conditions. Each scan was analyzed by generating standard *t*-test statistical significance maps and AMPLE relative activation maps. The stability of the activation maps in the motor regions of the cerebral cortex was compared for each map, both within each scan as a function of time and across different scans of the same subject. Simulated fMRI scans were also used to test the quantitative accuracy of signal detection using realistic distributions of known numbers of active voxels.

2.2. Subjects

Five healthy adult volunteer subjects (two females and three males; age, 19–50 years) were used in this study. All subjects were right-handed; all gave informed consent.

2.3. MR image acquisition

Scanning was done on both a 1.5- and a 4.0-T GE Signa LX scanner. Each subject was scanned at both field strengths. Gradient echo fMRI scans at 1.5 T were acquired using echo-planar imaging (EPI; TR/TE/flip=2 s/40 ms/90°, 22 slices, 5 mm thick) and spiral imaging (TR/TE/flip=2 s/40 ms/90°, 24 slices, 5 mm thick); scans at 4.0 T were spiral (TR/TE/flip=2 s/30 ms/90°, 24 slices, 3 or 5 mm thick). In each scan, images were acquired with 64×64 pixels and an FOV of 24 cm. A set of T2-weighted coplanar anatomical images (TR/TE/flip=3 s/84 ms/90°; FOV, 24 cm; 256×256 pixels) was acquired on each scanner for alignment to a whole-brain T1-weighted series acquired at 1.5 T (SPGR TR/TE/flip=24 ms/6 ms/30°; FOV, 25 cm; 1.5-mm slices; 256×256×128 pixels).

2.4. Motor paradigm

Subjects performed a hand-squeezing task to localize the motor cortex. Prompted by a visual cue, they alternated 9-s blocks of rest with 9-s blocks of opening and closing one hand (1/s). One full task cycle was 9 s of rest, 9 s of left-hand motion, 9 s of rest, 9 s of right-hand motion. Each

fMRI scan lasted 6 min 24 s (384 s); 10 full task cycles (360 s) were used in our analyses.

2.5. Analysis software

All data processing was performed using the fScan analysis program for fMRI [8] running on a Pentium IV computer running RedHat Linux. For this study, AMPLE voxel-counting algorithms and visualization tools were added to fScan's real-time processing options so that changes in dynamic thresholds and voxel counts could be monitored progressively as a function of scan time.

2.6. Defining regions of interest

Two large, rectangular 3-D regions of interest (ROIs) were defined for each subject. The ROIs covered the superior portion of each cortical hemisphere, approximately 7×7 cm in the axial plane and extending approximately 4 cm down from the top of the brain. Each ROI was drawn to be large enough to extend beyond the hand motor activation region on all sides but without overlapping with each other or with inferior active regions such as the cerebellum. These ROIs allowed the cortical activation for the left and right hands to be analyzed separately for each subject. Sample ROIs for one subject are shown in Fig. 1.

2.7. Generating statistical maps

Standard fMRI statistical maps were generated by fScan based on the timing of the task paradigm. For every scan, statistical activation maps were calculated based on *t*-test comparisons between task and control blocks, with a 4-s delay to compensate for hemodynamic latency. *t* values were calculated on a voxel-wise basis, with no explicit adjustment for multiple comparisons. No cluster size threshold was used.

No spatial smoothing was performed prior to calculating the statistical maps to preserve the spatial distribution of signals across neighboring voxels; for display purposes, the statistical maps themselves were spatially smoothed using a 4-mm smoothing kernel.

Normalized AMPLE maps of relative activation were created separately for each ROI. To do so, the voxel with the

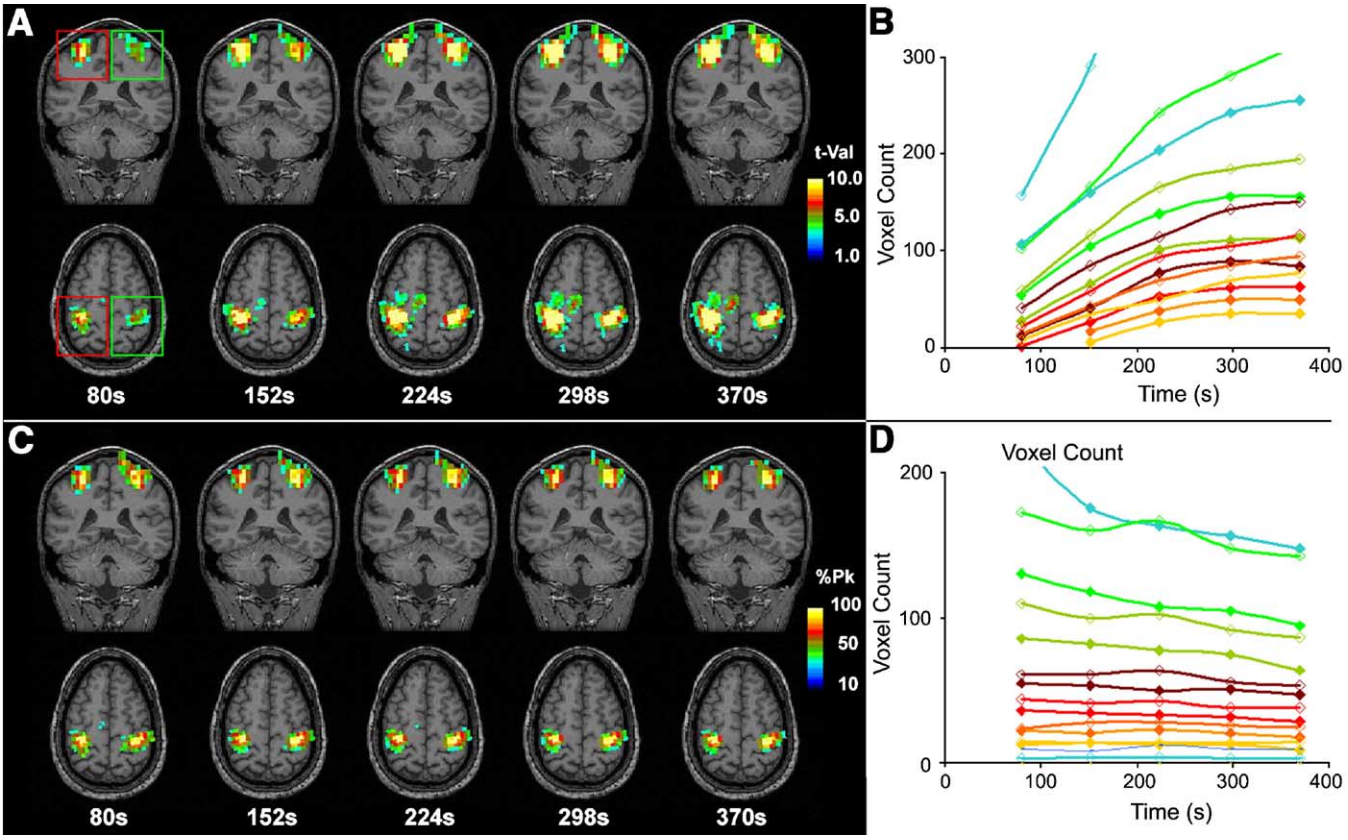

Fig. 1. A representative single-slice activation *t*-map and corresponding AMPLE map for one subject. Each panel shows the activation at five time points within a 6.5-min scan. At each time point, BOLD activity is shown superimposed on both coronal and axial T1 images at the level of the peak fMRI activation; the map is shown as a maximum intensity projection including all active voxels within 15 mm of the displayed slice plane to avoid slice selection bias. All images are shown in radiological convention. (A) Standard *t*-test maps showing all active voxels with *t* values ≥4.0. The rectangular ROIs used for counting voxels and AMPLE normalization in the right and left motor cortex regions are outlined in red and green, respectively, in the first pair of brain slices. (B) The number of active voxels as a function of scan time for seven *t*-value levels (3.0–9.0); the color code for *t*-value levels is the same as the color bar in Panel A. Open symbols are for the right-hemisphere ROI; closed symbols are for the left-hemisphere ROI. (C) The same data as in Panel A shown as AMPLE relative activation maps showing all active voxels ≥40% of the local peak *t*-value amplitude (active voxels were defined as *t* value ≥3.0). (D) The number of active voxels at each relative AMPLE (30–100% of peak *t* value); the color code for the curves is the same as the color bar in Panel C. The scans shown were acquired at 4.0 T using a spiral inward pulse sequence for sampling *k*-space.

highest statistical value in the ROI was first identified, and then, a new AMPLE statistical map was created, in which the value for each voxel was simply that voxel's t value converted to a percentage of the peak statistical value for the ROI (AMPLE value = t value/peak t value × 100).

2.8. Visualizing activation maps

In order to display and compare brain activation maps in a compact form, active fMRI voxels in each motor cortex ROI were displayed superimposed on T1-weighted anatomical images from the same subject. To avoid a slice selection bias, map values within 15 mm on either side of the image plane exhibiting peak motor activity were combined as maximum intensity projections to indicate the location and spatial extent of BOLD brain activity. In each subject, the same slice volumes were used for generating both standard and AMPLE fMRI maps.

2.9. Counting voxels as a function of statistical value

Within each ROI, the 3-D spatial extent of activation was calculated by counting the number of voxels with values at or above a minimum threshold level. For standard statistical maps, frequency distributions of voxel intensities were counted for nine uniformly spaced absolute t-value levels ($t \geq 2.0$ to $t \geq 10.0$). For AMPLE maps, frequency distributions of voxel values were counted for nine percentage levels (20–100%). For both t maps and AMPLE maps, voxels with absolute t values less than 2.0 were omitted from counting in order to avoid cluttering the frequency plots with large numbers of statistically insignificant voxels. All active voxels were counted at the end of every complete task cycle to measure temporal properties of the activation maps. The local peak height and both the absolute and relative frequency distributions were recalculated dynamically at every time point.

2.10. Computer simulations

Simulated fMRI image data sets were generated after the method of Skudlarski et al. [9] by adding 3-D clusters of "active" voxels with oscillating intensities of known frequency, amplitude and spatial distribution to human brain fMRI time-series images. Realistic clusters of simulated active voxels were first created by extracting clusters of active voxels detected in long (15 min) fMRI scans of healthy adult subjects performing the standard motor task described above or a simple block-designed reading task to activate language areas. Six clearly distinct and active clusters were identified from three different scans. For each cluster, a 3-D map was created from all voxels within 2 cm of the cluster center with a statistical t value of at least 1.0 over 450 image time points and two or more adjacent voxels with t values at least 1.0. Signal amplitude for each voxel was measured as the peak-to-peak intensity difference across an averaged cycle of the task time course divided by the mean image intensity for that voxel. Each sample cluster thus contained a distribution of active voxels with a range of signal amplitudes in a 3-D spatial pattern that matched the distribution actually observed in the test fMRI scans (a sample cluster is shown in Fig. 6). In the six sample clusters analyzed, the median signal amplitude in active voxels was between 0.9% and 1.2% of mean signal intensity, and peak amplitudes varied from 2.5% to 4.5%.

Simulated fMRI data sets were created by first converting the sampled cluster amplitude distributions into simulated active voxels by generating a time series of images in which all voxels in the cluster oscillated at the same phase and frequency (16 image time points per cycle) but with the original relative distribution of signal amplitudes. These simulated active voxels were then combined with one of two background human image data sets, each containing 64×64×24×450 voxels of typical EPI fMRI images. These background data sets had no significant time-varying intrinsic signal detectable with the 16-image oscillation period used for the simulated activations. For each simulation run, one of the sample clusters of active voxels was positioned over a brain region in one of the background EPI scans. Eight widely separated locations of the cerebral cortex were used for each cluster map/background combination. Any active cluster voxel that fell on a nonbrain voxel (i.e., image intensity <50% of the mean intensity of brain voxels) was removed from the cluster. The amplitude of signal oscillation for each active voxel was scaled by the mean intensity of the underlying background voxel to make the simulated BOLD signal have the same proportional amplitude (oscillation amplitude/mean intensity) as the original fMRI cluster. The oscillation amplitude of each active voxel was then reduced by 40% in order to compensate for the fact that the simulated voxels had more regular sinusoidal oscillations than the original data; this 40% reduction was empirically found to result in statistical activation maps of comparable significance to the original fMRI scans. Once scaled, the oscillating simulated BOLD cluster images were added to the background image time series to create new fMRI data sets with realistic spatial and temporal activity patterns. The median amplitude of simulated active voxels was 0.6%, and the peak amplitudes averaged 3% of mean signal intensity. The simulated image data sets were then analyzed using the same fMRI analysis methods described above for the real human fMRI data.

2.11. Generating receiver operator characteristic curves

Activation t maps and AMPLE maps generated for the simulated fMRI data sets were further analyzed by comparing the distribution of active voxels detected to the known distribution of true active voxels in each data set. Receiver operator characteristic (ROC) curves were generated for every simulated map [9]. To do so, for each threshold level (t value or relative t value) in each statistical map, every voxel was tested to see if it was detected as active in the map (i.e., at or above the counting threshold) and to see if it was truly active (i.e., had a 16-image oscillation amplitude above some input threshold). For

standard *t* maps, separate counts were generated for each *t*-value activation threshold, using a range of different input oscillation amplitude thresholds, ranging from 0.5% to 2.0% of mean signal intensity. For AMPLE maps, only one input signal amplitude threshold was used, but it was set to be the same percentage of true input signal as the AMPLE counting threshold. For example, using the 50% AMPLE threshold, voxels were tested to see whether their fMRI *t* value was at least 50% of the current peak *t* value and whether their input signal oscillation was at least 50% of the peak oscillation amplitude of all active voxels. Based on these comparisons, each voxel above the counting threshold was classified as true-positive or false-positive depending on whether it had a nonzero input oscillation amplitude. Similarly, voxels below the counting threshold were classified as either false-negatives or true-negatives depend-

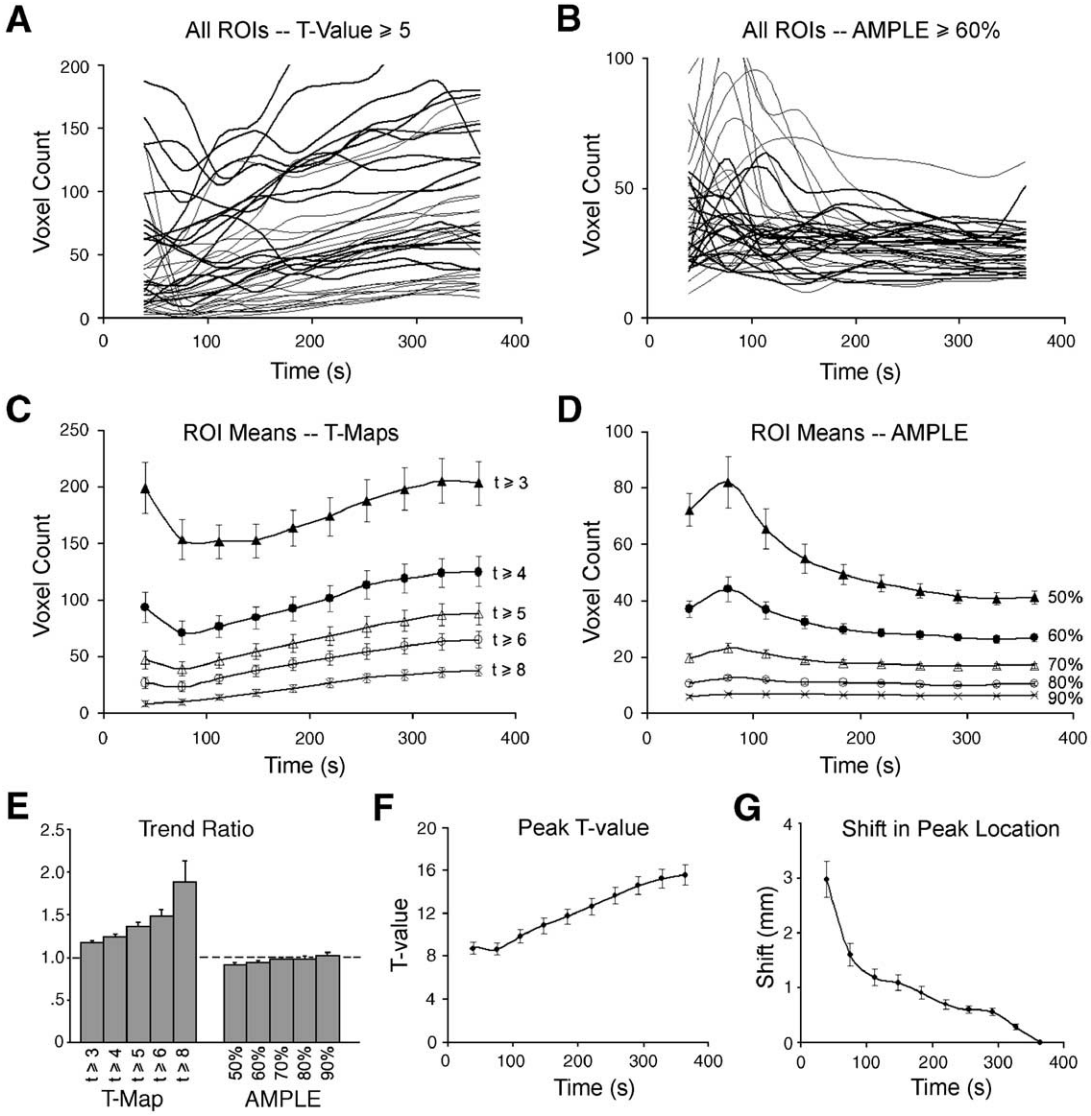

Fig. 2. Quantitative comparison of active voxel counts in standard *t* maps and AMPLE *t* maps as a function of scan time for all 40 motor ROIs in 20 scan runs across five subjects. (A) The number of active voxels using a *t*-value threshold of 5.0 is plotted separately for each ROI in every scan in the study. Scans performed at 1.5 T are shown with a thin gray line; scans at 4 T are plotted with a thick black line. For every scan, the number of active voxels above the fixed *t*-value threshold increased across time. (B) The same data as in Panel A, but this time, plotting voxel counts at the AMPLE 60% of peak level. Counts were initially variable but then stabilized for most scans. (C) The mean number of active voxels in standard *t* maps for all 40 scans is plotted as a function of scan time for five different *t* thresholds (labels shown on graph). (D) Mean counts of active voxels in AMPLE maps for all 40 scans plotted for five relative *t*-value levels (50–90% of the peak *t* value as shown). (E) Histogram showing the relative stability of AMPLE counts of active voxels, as compared with standard *t* maps. A "trend ratio" was calculated for each counting level (*t* threshold or AMPLE percentage) for every scan as the ratio of the number of counted voxels averaged over the last three task cycles divided by average counts over the previous three task cycles. For standard *t* maps, these ratios were significantly greater than 1 (increasing) for each threshold counted; for AMPLE maps, the ratios were approximately 1 (stable) for the five upper levels shown. (F) A plot of mean peak *t* value for all 40 scans showing a gradual increase in *t*-value levels as a function of scan time. (G) A plot of mean motion of the location of the statistical peak (in millimeters relative to the final position) for all 40 scans as a function of scan time. After 2 min, the 3-D position of the mean peak location shifted by less than 1 mm for almost every scan. Error bars in Panels C–G are standard errors of the mean.

ing whether their input oscillation amplitudes were above the input threshold or zero, respectively. Voxels that had input amplitudes less than the input threshold but greater than zero were ignored in ROC counts because misdetection of such voxels should not properly be considered an error. To avoid binning errors caused by relative activation values not being exactly the same as relative oscillation amplitudes, AMPLE voxels that were below the counting threshold but had an oscillation amplitude above the input threshold were also not considered false-negative errors if their t value was above the next lower counting threshold. For example, a simulated active voxel that oscillated at 53% of the peak oscillation amplitude but resulted in an activation t value that was 48% of the peak t value detected would not be classified as false-negative at the 50% counting threshold because it was above the next lower level (40%); however, it would be counted as false-negative at 50% if its t value were less than 40% of the peak t value.

Once voxels had been classified over the entire range of counting thresholds, an ROC curve was generated for each map by plotting the true-positive fraction (TPF) versus false-positive fraction (FPF) across all thresholds, as:

$$\text{TruePositiveFraction} = \text{TruePositives}/(\text{TruePositives} + \text{FalseNegatives})$$

$$\text{FalsePositiveFraction} = \text{FalsePositives}/(\text{TruePositives} + \text{FalsePositives})$$

An index score was also calculated for each ROC curve for comparison purposes. The index was calculated as the fractional area under the ROC curve for FPFs between 0 and 0.1 [9].

3. Results

3.1. Stability of fMRI maps as a function of scan time

The stability of fMRI activation across time for 6 min of a bimanual motor task is shown in Fig. 1 for a representative scan from one subject. The fMRI images in Fig. 1A show standard t maps, using a t-value threshold of 4.0, measured at different time points during the motor task. The color code for active voxels shows the distribution of t-value signal amplitude within each ROI (the ROIs are displayed in the first pair of images). Using a fixed t-value threshold, the number of brain voxels above threshold in each ROI increased progressively with repeated task cycles. This increase is quantified in Fig. 1B, which shows that at each statistical activation threshold, the number of active voxels progressively increased throughout the scan.

When brain activations were converted to AMPLE maps, however, the spatial extent of activation was stable for all AMPLE levels for almost the entire scan. This is illustrated in Fig. 1C and D in which the original t-map data in Fig. 1A and B were normalized as AMPLE contour levels, where each contour level represents a fixed percentage of the local peak t value. Fig. 1C shows all active AMPLE voxels with relative t values at least 40% of the peak value. The color code for active voxels shows the distribution of relative signal amplitude within each ROI. The spatial extent of AMPLE map voxels fluctuated over the first 1.5 min of the scan but then changed very little over the remaining 5 min. This resulted in stable appearance of activation maps over time (Fig. 1C) and stable numbers of active voxels at every percentage contour level (Fig. 1D).

Similar results were obtained in all 40 motor cortex ROIs measured in 20 scans acquired across the five test subjects, as shown in Fig. 2. Panels A and B show the results of all 40 ROIs individually for a t-value threshold of 5 and AMPLE threshold of 60%, respectively. Panels C and D combine the 40 ROIs into mean time plots of the number of active voxels over a range of different thresholds. In either case, when spatial extent of activation was quantified at a fixed t-value threshold, the number of active voxels increased progressively as a function of scan time. However, when the same active voxels were counted at relative AMPLE contour levels, the spatial extent of activation fluctuated initially but then stabilized at a plateau level. This difference in the slope of active voxel counts over time for the two mapping methods is quantified in Fig. 2E. Using standard t-value maps, the ratio of voxel counts observed in the last third of each run to the number detected in the middle third was significantly greater than 1.0, for each t-value threshold used for counting. In AMPLE maps, however, this ratio was

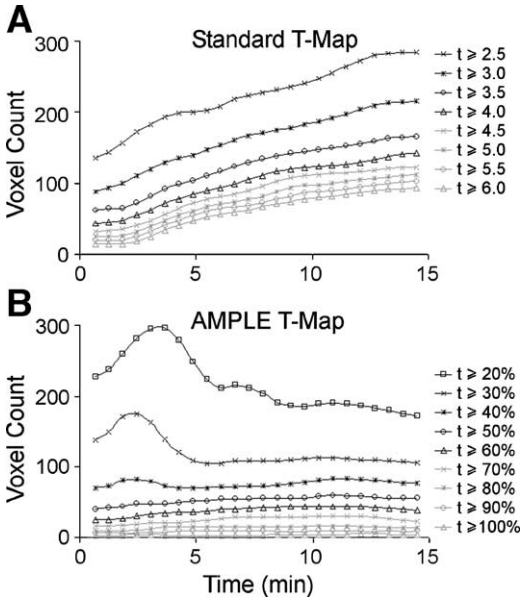

Fig. 3. Plots of active voxel counts across a range of counting thresholds for a representative 15-min scan of one subject performing the bimanual motor task. This long scan was performed at 1.5 T using the same behavioral task and scanning parameters as for the other EPI scans in this study. Panels A and B show time plots of the number of active voxels detected for different counting thresholds in standard t maps and AMPLE t maps, respectively.

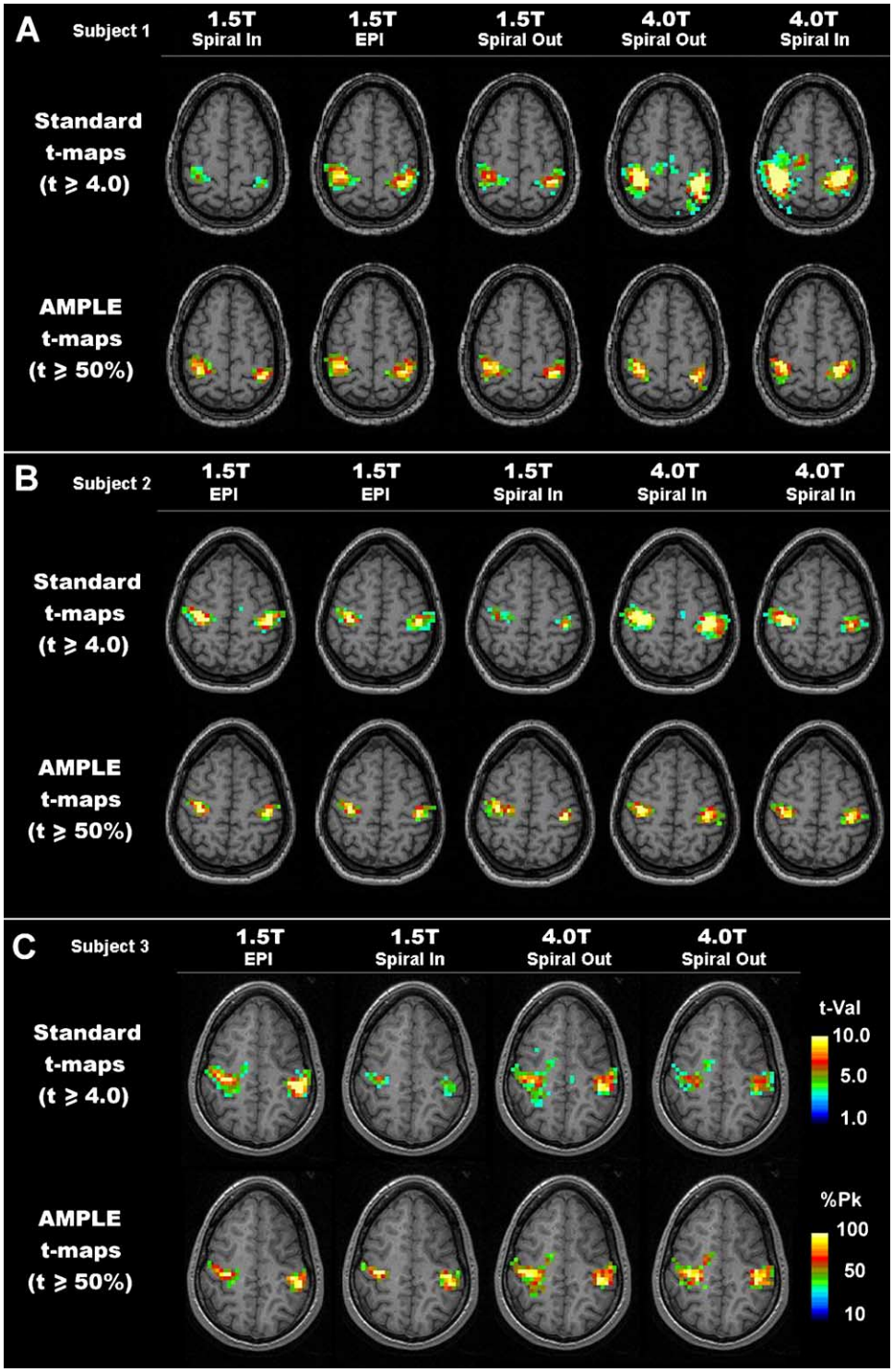

Fig. 4. Reproducibility across different runs, pulse sequences and field strengths. Each panel shows all runs acquired for a single subject performing the standard hand motor task. The top row in each panel shows maximum intensity projections of standard *t*-value maps (*t* threshold, 4.0), and the bottom row shows the AMPLE *t* maps for the same scans (relative *t*-value threshold 50% of peak *t*-amplitude). The color bars in Panel C apply to all three subjects. (A) Five different scans are shown for Subject 1 (the rightmost scan was used as the time-series example in Fig. 1). (B) Five scans acquired for Subject 2. In the second and fifth scans (counting from the left), the subject was instructed to perform the task more slowly and weakly than normal. (C) All four scans are shown for Subject 3. In the rightmost scan, the subject was instructed to perform the task slowly and weakly. In all three subjects, the distribution of color-coded voxels near the peak of activation was much more consistent across AMPLE maps than standard *t* maps when compared across multiple scans.

not significantly different than 1.0 for the same data counted as a function of relative activation level.

To test the behavior of active voxel counts over longer scan times, two subjects performed the alternating-hands motor behavioral paradigm for 15-min fMRI scans. As for the 6.5-min scans shown above, the number of active voxels in each ROI increased progressively over time for each absolute t-value statistical significance level, approaching plateau levels only for the lowest t-value thresholds ($t \leq 3.0$) after more than 10 min of scanning. In contrast, voxel counts at each AMPLE relative contour level reached plateau levels within the first couple of minutes and remained stable throughout the 15-min run (Fig. 3).

3.2. Stability of fMRI maps across different runs and at different field strengths

The reproducibility of the spatial distribution of brain voxels activated by a given behavioral task was tested by having subjects perform repeated runs of the same-hand motor task under a variety of imaging conditions. Each subject was scanned in two sessions, once in a 1.5-T scanner and once in a 4-T scanner, with images acquired using standard echo-planar and spiral fMRI pulse sequences. Fig. 4 summarizes the brain activation results for three representative subjects, showing both standard fMRI t maps and AMPLE relative t maps generated at the end of each scan. Although standard t-value maps showed BOLD fMRI brain activity in similar cortical locations across different runs for each subject, the activation level and spatial extent of statistically active voxels varied considerably. In particular, comparing standard t-value maps ($t \geq 4.0$) acquired at 1.5 T versus 4.0 T or acquired with EPI versus spiral imaging resulted in quite different-looking activation maps. However, when the same scans were converted to AMPLE relative t-value activation maps, the differences among imaging runs were much less. Thus, for AMPLE

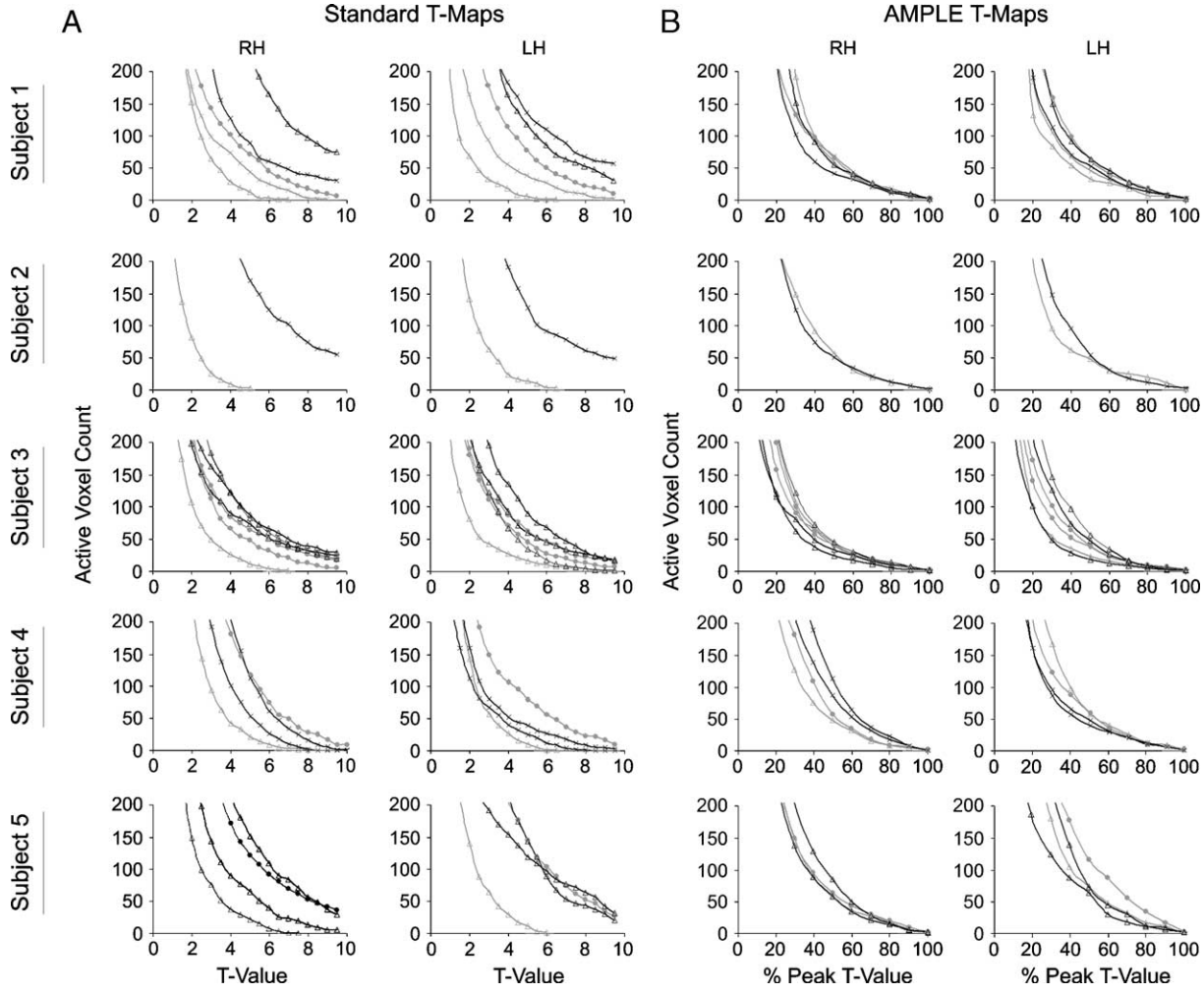

Fig. 5. Summary plots showing the distribution of active voxels counted in each brain region under different scanning conditions. Each graph plots all scans acquired for one ROI in one subject. Gray curves are for scans at 1.5 T; black curves are for scans at 4 T. Filled circles indicate EPI, open triangles are outward spiral scans and Xs are inward spiral scans. Voxel counts for scans obtained at 3 mm slice thickness were scaled by 60% for comparison with the 5-mm-thick scans. (A) Voxel counts are plotted as a function of standard t-value threshold for both ROIs in all five subjects. (B) The same data as in Panel A are replotted as voxel counts as a function of AMPLE relative peak levels.

levels that were at least 50% of the ROI's peak t value, the 3-D location and spatial extent of activation were qualitatively very similar from run to run, regardless of pulse sequence or scanner field strength used.

Quantitatively, the spatial distribution of fMRI brain activity detected in multiple scans of each subject was compared by counting the number of active voxels as a function of activity level (Fig. 5). For standard t maps, the number of active voxels detected at any particular t-value threshold varied widely within each subject. At a t value of 4, for example, there was an average variation of 900% across different scans in the number of active voxels detected in each motor ROI. For t values greater than 6, the variability was much greater in percentage terms because each subject had some scans with many voxels (50–100) whereas others had few to none. For AMPLE maps, in contrast, the number of active voxels detected at any relative activation level was relatively consistent across scans for each subject. The number of active voxels at or above 60% of the peak t value varied on average by only $21 \pm 5\%$ across all scanning conditions for each ROI active area.

Although all the results presented above were obtained using t-test statistical maps, the AMPLE approach also resulted in consistent activation maps across scan time and scan conditions when simple difference maps or correlation maps were computed for these motor scans. In the case of difference maps, there was little variation in the absolute amplitude of the difference signal across time; thus, the effect of converting to AMPLE maps was simply to normalize the MR signal amplitude to percentage values within each ROI (see Section 4). Standard correlation coefficient or z-score maps demonstrated variability over scan time and across scanning conditions; when converted to AMPLE maps, the relative distribution of active voxels in all maps stabilized within the first 2 min and was consistent across scanning conditions, both in location and number of active voxels (data not shown).

3.3. Simulation studies

Ninety-six simulated fMRI data sets were created with realistic amplitude and spatial distributions of active voxels as described in Section 2. Both standard t maps and AMPLE t maps were generated for each data set, and ROC curves

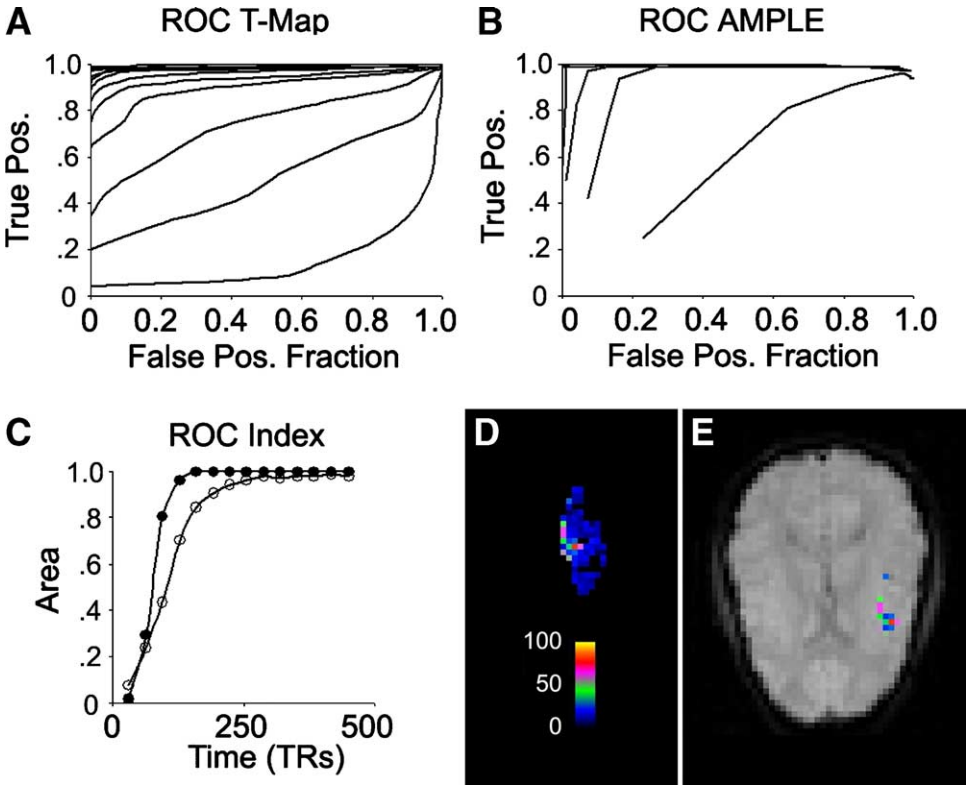

Fig. 6. ROC analysis of simulated fMRI scans. Families of mean ROC curves for all 96 simulated scans in which voxels were counted in standard t maps or AMPLE t maps are shown in Panels A and B, respectively. Each curve shows the ROC relationship in maps computed after a different number of task cycles; in both panels (A and B), the lowest curve was after two task cycles, with higher curves corresponding to progressively longer simulated scan times. The families of ROC curves for standard t maps (open circles) and AMPLE maps (closed circles) are compared in Panel C by plotting the area under each curve for FPFs less than 0.1 as a function of scan duration. In Panels A, B and C, values close to the upper left corner are better. Panel D shows an example of one of the sample active clusters used in the simulations; this cluster was extracted from a real motor fMRI scan as described in Section 2. Panel E shows an AMPLE activation map ($t \geq 30\%$) analysis of a simulated scan using the same cluster. The color scale in Panel D applies to both maps, as percent oscillating signal amplitude in Panel D and as percent peak t value detected in Panel E.

were calculated for each map to assess how effectively it detected that data set's input population of simulated voxels. Fig. 6 shows families of mean ROC curves generated as a function of scan duration, averaged across the 96 simulated data sets. For standard *t* maps, the ROC curves gradually shifted toward the upper left corner (optimal sensitivity) with increasing scan time. ROC curves for AMPLE maps of the same scans also shifted toward the upper left corner as a function of scan duration but more rapidly than for the standard *t* map. This difference is clearer in the ROC index plot, which shows that the area under the ROC curves reached its optimal value (1.0) for AMPLE maps in half the time it took for standard *t* maps.

We also used the simulated data sets to test the accuracy with which any particular map could distinguish the true population of active voxels (Fig. 7). For standard *t* maps, the simulated data sets resulted in gradually increasing numbers of active voxels detected at each *t*-value threshold as a function of time, as expected given the range of signal amplitudes present in the simulated input activation. Given this time-dependent increase in sensitivity to weakly activated voxels, therefore, any direct comparison of detected voxels to truly active voxels in standard *t* maps needed to specify which threshold to use for counting and what signal oscillation amplitude to define as truly active. For example, Fig. 7B and C shows the accuracy of true active voxel detection using standard *t* maps for the simulated data with *t*-value thresholds of 4 and 6, respectively. For each, a range of different input signal amplitudes was used, resulting in different numbers of oscillating voxels being considered truly active. The graphs show that with increasing numbers of task cycles, the number of detected voxels gradually moved toward the number of truly active voxels for each *t*-threshold/signal

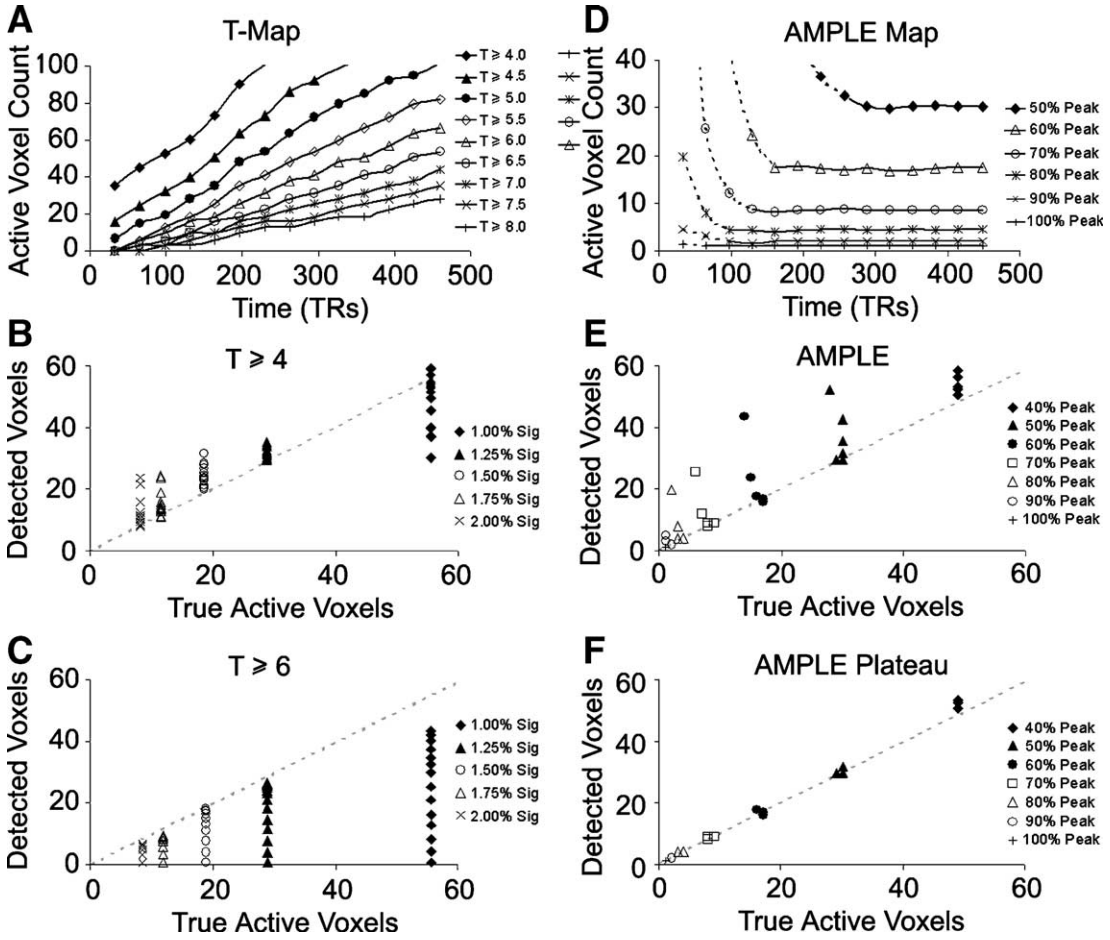

Fig. 7. Plots of voxel counts detected in simulated fMRI scans. Results for all 96 simulated scans are shown. (A) Mean voxel counts at different threshold levels in standard *t* maps as a function of scan time. (B) Comparison of the number of voxels detected in standard *t* maps using a *t* threshold of 4 to the true number of active voxels included in each simulated scan. Each symbol plots the results considering a different input signal oscillation amplitude as the criterion for truly active voxels (see text). The vertical columns of each symbol represent voxel counts after different numbers of task cycles. The symbols closest to the ideal detection dotted line correspond to the largest number of task cycles. (C) The same scans plotted as in Panel B but using a *t* threshold of 6. The stable regions of each curve (within 10% of the final value) have a solid black line, and the variable regions are shown with a dotted gray line. (D) The same scans as in Panel A plotted by different AMPLE relative thresholds as a function of time. (E) AMPLE voxel counts compared with true numbers of active voxels, in which the criterion for truly active voxels was the relative amplitude of the oscillating input signal (% peak) corresponding to each relative AMPLE level (see text). All AMPLE levels at all time points are shown in Panel E; longer times again correspond to the points closest to the ideal detection line. (F) The same data as in Panel E but excluding AMPLE voxel count data corresponding to the dotted gray line portions of Panel D.

amplitude condition. For any given *t*-value threshold, this was associated with increased sensitivity (TPF) over time; for lower *t*-value thresholds there was also an increase in specificity (decreased FPF) over time.

Counts of active voxels in AMPLE maps of the simulated data sets followed a different pattern. Just as for the real human subject data, the number of active voxels detected at each relative activation level initially rose rapidly and then fell as a function of scan time, gradually reaching stable plateau levels. The solid portions of the curves in Fig. 7D were regions in which voxel counts differed from the final time point value by less than 10%. Comparing detected voxel counts to truly active voxels in AMPLE maps was simpler than for standard *t*-map counts because the number of voxels in each AMPLE relative activation level could be compared with the number of simulated active input voxels at the corresponding relative oscillation amplitude level (see Section 2). As such, the graph in Fig. 7E summarizes the accuracy of signal detection over time for all AMPLE levels. As for the standard *t* maps, the number of detected voxels at each level gradually approached the true number of active voxels over time. When the same data were replotted, but omitting the preplateau time points at each AMPLE level (the dotted lines in Fig. 7D), the result was Fig. 7F. For these simulated scans, the active voxels detected in the stable plateau region of any relative AMPLE map level accurately reflected the true distribution of oscillating voxel signal intensities present in each data set.

4. Discussion

The normalized AMPLE maps introduced here are intended to augment standard fMRI statistical maps by providing a measure of the stability and reproducibility of the brain activation signal. The assumption underlying this approach is that a subject performing a behavioral task should produce a characteristic pattern of brain activity, which will elicit a characteristic spatial pattern of BOLD signal. The statistical significance of fluctuations across repeated behavioral trials can be used to detect regions of task-dependent BOLD signal, whereas changes in the relative distribution of the signal across neighboring voxels can be used to evaluate whether the spatial pattern is stable.

The results presented above demonstrate the validity of this approach. Standard *t*-test fMRI maps for a simple motor behavioral task detected motor cortex activation but varied considerably as a function of scan time and across different scan runs of the same subject. The relative spatial distribution of active voxels, as revealed in the AMPLE maps for those same scans, was highly reproducible, however. Both the location and number of active voxels at each AMPLE level (percentage of local peak amplitude) were consistent over time and across scans. The results indicate that within each active ROI, the amplitude of the statistical activation peak varied with the number of trials performed, the magnet strength and imaging parameters used, but the relative spatial pattern of the brain BOLD activation pattern was consistent for each task. Although all the scans performed in this study involved a simple motor task, we have obtained similar results using simple language mapping paradigms as well (unpublished data).

The AMPLE algorithm simply involves performing a spatially adaptive normalization in which values in each active region within a statistical activation map are converted to percentage of the local peak value. The motivation for locally adaptive normalization is based on the notion that different brain areas are involved in different aspects of behavior, and thus, the spatial distribution of voxel activity should generally be considered for each brain region independently. In the bilateral motor task, for example, adaptive normalization ensures that the AMPLE activation map peak for each hand's motor area is always 100%, even if the subject has quite different absolute levels of activation for each hand. Obviously, the choice of local ROIs depends on the application; in some cases, the entire brain may be considered a single ROI, whereas in others, it might be divided into smaller ROIs either manually, as in the motor cortex example described here, or automatically, using a cluster segmentation algorithm. Provided that the ROI is bigger than the active area of interest and does not contain multiple different activation peaks, we have found that the exact size or shape of the ROI does not significantly affect the AMPLE maps produced (not shown).

The AMPLE normalization method is designed to supplement and not replace traditional statistical significance fMRI mapping. Significance of activation is determined based on voxel *P* values in the statistical activation map (whether calculated as *t* test, correlation coefficient, *z* score, etc.), and significance thresholds are used to decide which voxels are active in any given scan. The advantage in adding normalized AMPLE maps to the analysis is that they help reveal both the temporal and spatial stability of the overall pattern of activation within each brain region. We have focused on *t*-test activation mapping in this report because this simple statistical measure combines the amplitude of the activation signal (difference in mean intensities) as well as the amplitude of the noise (variance). Relative maps of significance alone (e.g., *P* values or total variance maps) provide little information about the relative distribution of brain activity signal amplitude. Conversely, simple difference maps of task-dependent intensity variations tend to be fairly stable as a function of scan duration, but on their own they provide very little information about statistical significance. A key feature of the AMPLE approach, therefore, is the combination of statistical significance and relative levels of activation signals in order to assess stability of fMRI maps.

Measuring the stability of statistical activation maps is important in three different ways. First, the fact that the distribution of active voxels at each relative statistical activation level stabilizes over time means that such

measurements can provide a valuable quality assessment tool for determining whether enough trials have been collected to provide a reliable map. Once the number of active voxels at a particular AMPLE contour level reaches a plateau value, our results indicate that continued scanning is not likely to provide much additional information (at that level). Our simulation results suggest that stabilization of the number of voxels detected at a relative statistical significance level indicates that all truly active voxels at that relative activation level have been correctly detected.

Conversely, if the distribution of active voxels at different AMPLE contour levels does not plateau, it suggests that the subject may be having problems performing the task or that more scanning may be needed before one can have confidence in that fMRI map. By implementing this adaptive analysis within real-time fMRI analysis software, the gradual stabilization of adaptive voxel counts offers a simple and robust quality assurance measure for real-time evaluation of data acquisition. Such quality assurance tools can be a significant practical advantage when scanning difficult subject populations such as children or the elderly or when performing clinical fMRI scans.

The second important advantage in generating normalized AMPLE maps is that such maps provide a meaningful way to compare the same aspects of brain activation across different scans. Unlike a standard fMRI statistical activation map, which needs to be qualified as having been measured on a particular scanner, with a particular pulse sequence and after a particular number of trials, our results show that AMPLE maps have the advantage of being quite stable across different scans. They convert any suprathreshold activation map to a standardized set of percentage contour levels; voxels that are above threshold in different scans will tend to appear in the same color-coded contour level in each AMPLE map, while differences in overall activation strengths will be reflected in differences in how many of the low-percentage contour levels appear in each map. Different AMPLE maps should be directly comparable across scans or across subjects, in terms of the location and spatial extent of activation at each contour level. As BOLD fMRI is an indirect measure of brain function in which the vascular signal spreads well beyond the site of neuronal activity, active areas defined by statistical significance alone are almost certain to include some brain voxels that are not neuronally involved in the task. Some studies have reported that only the most significant voxels within an fMRI active brain area may actually be closely correlated with electrophysiologically measured brain activity [10,11]. AMPLE normalization provides a simple tool for identifying reproducible spatial boundaries of the BOLD signal for each activation as stable normalized contours of its own local distribution pattern. These boundaries should provide a more reliable basis for comparing spatial activation patterns than simply depending on absolute measures of statistical significance of the BOLD signal.

The third advantage of using normalized fMRI activation maps is that the normalization process itself can be used as a means of calibrating differences in order to combine data obtained at different times or scan sites or for performing meta-analysis of fMRI data collected under different conditions. If differences among scans can be largely reduced by normalization, the amount of scaling involved in the normalization process could provide global calibration metrics that might help compare other scans acquired under similar conditions. Thus, even if a particular study is interested in comparing absolute BOLD activation levels across multiple scans, calibration parameters derived from AMPLE normalization scans could help make such comparisons more meaningful. Such normalization metrics are also well suited for use in longitudinal studies that might otherwise be jeopardized by unanticipated changes in scanner software or hardware. As the emphasis in brain imaging shifts toward increased data sharing and combining imaging data across multiple research sites, techniques for calibrating differences among scanners and among imaging methods will become increasingly important [6].

In summary, the AMPLE normalization approach demonstrates that fMRI can be used to produce reproducible maps of the spatial distribution of task-dependent BOLD activity under a wide range of scanning conditions. In this study, we have focused on normalization of *t*-test maps derived from a simple block-design behavioral task, but the same principle works for other types of statistical maps and could be applied to event-designed fMRI studies as well. By considering the statistical activation of each voxel relative to the activation of its neighbors, the simple and robust AMPLE method provides a useful additional tool for evaluating and comparing reproducibility of brain activations across scans in a wide range of fMRI applications.

Acknowledgements

The author thanks Imran Deshmukh, Katy Harris and Jimmy Dias for their help and many useful discussions. This work was supported by PHS Grants NS37746 and NS41328.

References

[1] Bandettini PA, Wong EC, Hinks RS, Tikofsky RS, Hyde JS. Time course EPI of human brain function during task activation. Magn Reson Med 1992;25:390–7.

[2] Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, et al. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. PNAS 1992;89:5675–9.

[3] Heeger DJ, Ress D. What does fMRI tell us about neuronal activity? Nat Rev 2002;3:142–51.

[4] Heuttel SA, Song AW, McCarthy GM. Functional magnetic resonance imaging. Sunderland (MA): Sinauer; 2004.

[5] Liu JZ, Zhang L, Brown RW, Yue GH. Reproducibility of fMRI at 1.5 T in a strictly controlled motor task. Magn Reson Med 2004;52:751–60.

[6] Zou KH, Greve DN, Wang M, Pieper SD, Warfield SK, White NS, et al. Factors impacting the reproducibility of functional MR imaging: preliminary results of a prospective multi-institutional study by the biomedical informatics research network. Radiology 2005;237:781–9.

[7] Vlieger EJ, Majoie CB, Leenstra S, den Heeten GJ. Functional magnetic resonance imaging for neurosurgical planning in neuro-oncology. Eur Radiol 2004;14:1143–53.

[8] Voyvodic JT. Real-time fMRI paradigm control, physiology, and behavior combined with near real-time statistical analysis. NeuroImage 1999;10:91–106.

[9] Skudlarski P, Constable TR, Gore JC. ROC analysis of statistical methods used in functional MRI: individual subjects. NeuroImage 1999;9:311–29.

[10] Roux FE, Ibarrola D, Tremoulet M, Lazorthes Y, Henry P, Sol JC, et al. Methodological and technical issues for integrating functional magnetic resonance imaging data in neuronavigational systems. Neurosurgery 2001;49:1145–57.

[11] Arthurs OJ, Boniface SJ. What aspect of the fMRI BOLD signal best reflects the underlying electrophysiology in human somatosensory cortex? Clin Neurophysiol 2003;114:1203–9.