ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 32 (2006) 1656–1668

Reducing inter-scanner variability of activation in a multicenter fMRI study: Role of smoothness equalization

Lee Friedman,[a,*] Gary H. Glover,[b] Diana Krenz,[c] and Vince Magnotta[d]
The FIRST BIRN[1]

[a]Department of Psychiatry and Human Behavior, University of California-Irvine, CA 92617, USA
[b]Department of Radiology, Stanford University, Stanford, CA 94305, USA
[c]The MIND Institute, Albuquerque, NM 87106, USA
[d]Department of Radiology, University of Iowa, Iowa City, IA 52242, USA

Received 11 February 2005; revised 7 December 2005; accepted 30 March 2006
Available online 27 July 2006

Scanner-to-scanner variability of activation in multicenter fMRI studies is often considered undesirable. The purpose of this investigation was to evaluate the effect of a new procedure, "smoothness equalization", on reducing scanner differences in activation effect size as part of a multicenter fMRI project (FIRST BIRN). Five subjects were sent to 9 centers (10 scanners) and scanned on 2 consecutive days using a sensorimotor fMRI protocol. High-field (4 T and 3 T) and low-field (1.5 T) scanners from three vendors (GE, Siemens, and Picker) were included. The activation effect size of the scanners for the detection of neural activation during a sensorimotor task was evaluated as the percent of temporal variance accounted for by our model (percent of variance accounted for or PVAF). Marked scanner effects were noted for both PVAF as well as the degree of smoothness of the raw and processed images. After smoothness equalization, there was a dramatic (low field) or consistent (high-field) reduction in scanner-to-scanner variation of activation. It was shown that the likely basis of the scanner differences in smoothness was differences in k-space filtering algorithms. This work highlights the need to account for differences in smoothness when comparing scanners on activation effect size in multicenter fMRI studies.

Introduction

Multicenter brain imaging studies have a number of advantages and are becoming increasingly common. The most obvious advantage of a multicenter study is the ability to accumulate large samples of subjects from potentially diverse demographic distributions. This is useful when studying rare diseases, but even when this is not the case, a large group of subjects can be especially useful to increase the statistical power needed for more sophisticated multivariate statistical analyses (e.g., structural equation modeling) or data mining. A large sample of subjects can also be used to elucidate and confirm small, but important, effects. Another way in which multicenter studies can be used is to study scanner differences that are of interest and importance. Such scanner differences can suggest explanations for disparate published results and can thus enlighten researchers as to the technical factors that affect their study. Finally, multicenter studies tend to engender the interaction and collaboration of scientists with different views. Discussions in the planning and execution phases of multicenter studies can serve to bridge differences and to promote the development of consensus views. For these reasons, there have been a number of recent multicenter brain imaging studies in several domains: structural brain imaging (de Certaines et al., 1993; Van Haren et al., 2003; Schnack et al., 2004), magnetic resonance spectroscopy (Paley et al., 1996), magnetization transfer (Silver et al., 1999; Berry et al., 1999) and fMRI (Casey et al., 1998).

The Biomedical Informatics Research Network (BIRN) is a National Center for Research Resources initiative that fosters distributed collaborations in biomedical science by utilizing information technology innovations (http://www.nbirn.net). The BIRN currently consists of three "test bed" projects that are conducting structural and functional studies of brain disease: The FIRST BIRN (Functional Imaging Research Schizophrenia Testbed BIRN), or fBIRN testbed, from which the present report emanates, uses multicenter fMRI to study regional brain dysfunction related to the progression and treatment of schizophrenia. The Morphometry BIRN testbed uses sophisticated structural MR techniques and tools to examine unipolar depression, mild Alzheimer's disease and mild cognitive impairment. The Mouse BIRN project studies animal models of multiple sclerosis,

* Corresponding author. 1312 Michael Hughes Dr. NE, Albuquerque, NM 87112, USA.
E-mail address: lfriedman10@comcast.net (L. Friedman).
[1] Functional Imaging Research Schizophrenia Testbed-Biomedical Informatics Research Network, http://www.nbirn.net, NIH-NCRR, Bethesda, MD 87131, USA.
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2006.03.062

schizophrenia, Parkinson's disease, ADHD, Tourette's disorder, and brain cancer.

The key challenge for multicenter imaging studies stems from the fact that centers differ in numerous ways, and some of these differences can have powerful effects on the imaging results. There are two global strategies to deal with variance due to scanner: treat it as random variance, or try to understand and reduce it. In this investigation, the later approach is emphasized. The initial goal of the present effort was to evaluate scanner differences in the sensitivity to the BOLD effect in fMRI. The measure of sensitivity we used was the activation effect size, explained below. As it became clear that there were important differences in activation effect size that were not simply a reflection of field strength, we began to look for other factors that might explain such differences.

In our analyses of these multiscanner data, in addition to analysis of activation effect size, we emphasized the analysis of spatial smoothness. Increased spatial smoothness increases temporal SNR in fMRI and can greatly increase the ability to detect BOLD signal changes (Parrish et al., 2000), but this increase in SNR comes at the price of decreased spatial resolution. The effect is quite strong, even when the image smoothness results from varying degrees of k-space filtering (Lowe and Sorenson, 1997). Furthermore, Strother and colleagues have repeatedly shown that smoothing fMRI data can markedly enhance the reproducibility of fMRI results collected from a single scanner (LaConte et al., 2003; Strother et al., 2004). Since the usefulness of the multicenter strategy depends on the reproducibility of fMRI results across scanners, it may be important to pay close attention to variations in spatial smoothness produced by scanners with different field strengths and from different vendors. It is paramount to keep in mind throughout our report that when we say "increased smoothness", this can be rephrased as "decreased spatial resolution" and vice versa.

A previous report on this data has been published by Zou et al. (2005). Zou et al. (2005) assessed the effect of several factors on the reproducibility of fMRI activations, using a novel method to develop a gold standard activation map. These authors reported that 3 T scanners had higher CNR than 1.5 T scanners and also that results from 3 T scanners were more reproducible than results from 1.5 T scanners. The present study is focused more on understanding the role that smoothness plays in explaining scanner differences activation effect size.

In the present study, we compare the 10 FIRST-BIRN scanners (Table 1) on activation effect size and spatial smoothness. Both high (3 T and 4 T) and low-field (1.5 T) scanners were represented, as well as scanners from Siemens (Siemens Medical Solutions of Siemens AG, Malvern, Pennsylvania, USA), GE (GE Healthcare Technologies, Waukesha, Wisconsin, USA) and Picker (Royal Philips Electronics, Amsterdam, The Netherlands). The fBIRN Phase I "Human Phantom" study, in which 5 subjects were scanned on all 10 scanners, provided a unique opportunity to compare scanners on activation effect size and image smoothness. The results illustrate marked scanner differences and suggest methods to reduce these differences and thus enhance multicenter reproducibility.

Materials and methods

Subjects

Five healthy, English-speaking males (mean age: 25.2, range = 20.2 to 29) participated in this study. All were right-handed, had no history of psychiatric or neurological illnesses and had normal hearing in both ears. Each subject traveled to 9 sites (10 scanners) (Table 1), where they were scanned twice over a period of 2 days for a total of 20 scans per participant. There were no missing data, i.e., all 100 scans (5 subjects × 2 Visits × 10 scanners) were available for analysis. All subjects were instructed to avoid alcohol the night before the study, caffeine 2 h prior to the study and to get a normal night's sleep the night before a scan session. Informed consent was obtained from every subject before participation in this study and before every scan session. The approval of every site's International Review Board was obtained before conducting this study.

Image acquisition

A bite bar was used to stabilize each subject's head and was placed in the subject's mouth at the beginning of each scan session. An initial T2-weighted, anatomical volume for functional overlay was acquired for each subject (fast spin-echo, turbo factor = 12 or 13, orientation: parallel to the AC–PC line, number of slices = 35, slice thickness = 4 mm, no gap, TR = 4000 ms, TE = approximately 68, FOV = 22 cm, matrix = 256 × 192, voxel dimensions = 0.86 mm × 0.86 mm × 4 mm). The parameters for this T2 overlay scan were allowed to vary slightly from scanner to scanner according to field strength or other local technical factors. The anatomical scan was followed either by a working memory task (total of 14.7 min) or an attention task (total of 16 min). Three subjects always performed the working memory task and two subjects always performed the attention task. Over the next hour

Table 1
Description of hardware and sequences of the nine sites (10 scanners) participating in this study, five 1.5 T scanners, four 3 T scanners, and one 4 T scanner

Center	Abbreviation	Field strength	Manufacturer	RF coil type	Functional sequence
Brigham and Women's	BWHM	3.0 T	GE	GE TR Research Coil	EPI
Duke/UNC	D40T	4.0 T	GE Nvi LX	TR quadrature head	Spiral
Duke/UNC	D15T	1.5 T	GE Nvi LX	TR quadrature head	Spiral
University of Iowa	IOWA	1.5 T	GE Signa CV/i	TR quadrature head	EPI
Mass. General Hospital	MAGH	3.0 T	Siemens Symphony Trio	TR quadrature head	EPI-Dual Echo
University of Minnesota	MINN	3.0 T	Siemens Symphony Trio	TR quadrature head	EPI
University of New Mexico	NMEX	1.5 T	Siemens Sonata	RO quadrature head	EPI
Stanford University	STAN	3.0 T	GE CV/NVi	Elliptical quadrature head	Spiral in/out
University of California, Irvine	UCIR	1.5 T	Philips/Picker	RO quadrature head	EPI
University of California, San Diego	UCSD	1.5 T	Siemens Symphony	TR quadrature head	EPI

or so, subjects performed 4 runs of a sensorimotor task (described below, 4.25 min per run, total of 17 min), 2 runs of a breath-hold task (4.25 min), and 2 resting state scans (4.25 min). These 8 scans were performed in a counterbalanced order. The entire scan session was repeated the following day. In the present report, only data from the sensorimotor task will be presented.

The functional data were collected using echo-planar (EPI) trajectories (7 scanners) or spiral trajectories (3 scanners) (see Table 1) (orientation: parallel to the AC–PC line, number of slices = 35, slice thickness = 4 mm, no gap, TR = 3000 ms, TE = 30 ms on the 3 T and 4 T scanners, 40 ms on the 1.5 T scanners, FOV = 22 cm, matrix = 64 × 64, voxel dimensions = 3.4375 mm × 3.4375 mm × 4 mm). MAGH (usually abbreviated MGH, see table for scanner abbreviations) employed a double echo EPI sequence. All the spiral acquisitions were collected on General Electric (GE) scanners. The sensorimotor task produced 4 runs of 85 volumes each (85 TRs). The RF coils used varied with each scanner (Table 1).

Sensorimotor task

The sensorimotor task was designed by one of the authors (GHG) initially for other calibration purposes and employed a block design, with each block taking 10 TRs (30 s) beginning with 5 TRs (15 s) of rest (subject instructed to stare at fixation cross) and 5 TRs (15 s) of sensorimotor activity (see below). There were 8 full cycles of this followed by a 5 TR rest period at the end for a total of 85 TRs (4.25 min). During the active phase, subjects were instructed to tap their fingers bilaterally in synchrony with binaural tones, while watching an alternating contrast checkerboard. The checkerboard flash and tone presentation were simultaneous. The subjects were instructed to tap their fingers in an alternating finger-tapping pattern (index, middle, ring, little, little, ring, middle, index, index…) in synchrony with the tones and checkerboard flashes. The thumb was not used in this study. Each tone was 166 ms long with 167 ms of silence. The tone sequence has been described as annoying and atonal (Midi notes 60, 64, 72, 76, 80, 84, 88, 86, 82, 78, 74, 70, 66, 62, 58). The subject's responses were recorded and monitored with the PST Serial Response Box (Psychology Software Tools, Inc., Pittsburgh, PA) (except for NMEX, which used a more ergonomically designed, custom built device).

Data analysis

The first step of image processing was accomplished using Analysis of Functional NeuroImage (AFNI) software (Cox, 1996). All large spikes in the data were removed from each sensorimotor run, and each run was motion corrected (i.e., spatially registered to the first volume of the run). The data were then slice-time corrected. A mean functional (T2*) image and also a residualized timeseries was created by detrending each run with a 2nd order polynomial. The mean T2* image for each run was spatially normalized to an EPI canonical image in MNI space using tools available in SPM5b (http://www.fil.ion.ucl.ac.uk/spm/). This included affine transformations and 3 non-linear iterations. The spatial transformations were applied to the detrended timeseries data as well, and the timeseries was resampled at a 4 × 4 × 4-mm voxel size. In the baseline case, the mean T2* image and the detrended data were summed prior to statistical analysis.

However, for smoothness equalization (see below), the residualized timeseries was first smoothed to a target smoothness (7 mm FWHM) to remove scanner-related smoothness differences. These smoothed residuals were then added back to the mean EPI image prior to level 1 statistical analysis.

For level 1 statistical image analysis, Keith Worsley's package, FMRISTAT (Worsley et al., 2002; Liao et al., 2002) (http://www.math.mcgill.ca/keith/fmristat/) was employed. The first 2 volumes were discarded to allow for T1 saturation effects to stabilize. The statistical analysis was based on a linear model with correlated errors. The design matrix of the linear model, consisting of the off and on periods of the blocks, was first convolved with a hemodynamic response function modeled as a single gamma function (time to peak 4.7 s, FWHM = 3.8 s, Cohen, 1997) to produce the main regressor. In addition, the first derivative of this regressor was also included in the model to allow for some adjustment of the timing and shape of the response (Friston et al., 1998). Temporal drift was removed by adding a linear and quadratic contrast to the model. The correlation structure was modeled as a 1st degree autoregressive process. At each voxel, the autocorrelation parameters were estimated from the least squares residuals using the Yule-Walker equations, after a bias correction for correlations induced by the linear model. The autocorrelation parameters were first regularized by spatial smoothing, then used to 'whiten' the data and the design matrix. The linear model was then reestimated using least squares on the whitened data. The parameter passed to level 2 statistical analysis was a common measure of effect size, i.e., the percent of variance accounted for (PVAF) defined as 100*(sum of squares to due the main and derivative regressors)/(total sum of squares). The numerator can be thought of as variance due to the model and the denominator can be thought of as total temporal variance about the mean. Slight modifications to FMRISTAT were required to produce this measure, which will be referred to also as "activation effect size". The PVAF was given the sign of the beta weight for the main (i.e., non-derivative) regressor, although negative PVAF measures were generally not observed in this study.

Creation of ROIs

The sensorimotor task was expected to activate motor areas related to finger-tapping and sensory areas related to hearing tones and observing the flashing checkerboard. We developed a set of "functionally defined" ROIs (Fig. 1) based on an analysis of the data in the present study. First, we performed a series of one-sample *t*-tests for each scanner and for each subject. The ROIs include only voxels that were activated at all 10 scanners with an uncorrected p value <0.00001 and in all 5 subjects with an uncorrected p value <0.00001. There were 8 ROIs identified: left and right motor cortex, left and right auditory cortex, left and right cerebellar cortex, bilateral visual cortex and the supplementary motor area (Fig. 1).

Measuring scanner smoothness

Smoothness was estimated using the AFNI program 3dFWHM, which implements the method suggested by Forman et al. (1995). The approach assumes that the data were originally spatially independent, and it then proceeds to estimate the Gaussian kernel width (FWHM, in mm) of a filter that would have produced data matching the smoothness of the image in question. Generally, the estimate was made on all volumes (and averaged within a run) after each run had been detrended with a 2nd order polynomial. With

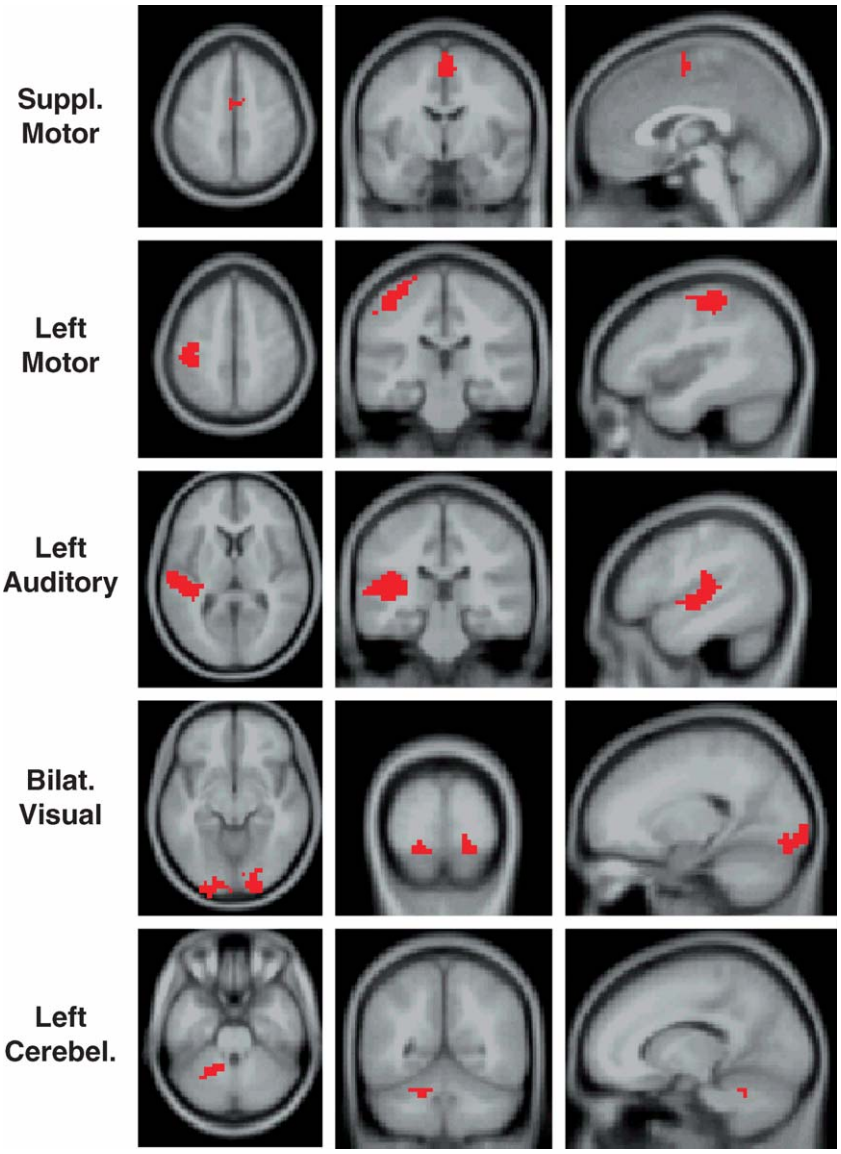

Fig. 1. Five of 8 ROIs employed in this study. For each ROI, an axial, coronal and sagittal view are presented. The remaining 3 ROIs (right motor, right auditory, and right cerebellum) were comparable to the contralateral ROIs shown in this figure.

both the trend and the constant removed from these images, these zero-centered (residual) images contained no brain structure information and the contribution of brain structure to smoothness was thus removed. The AFNI measure provides estimates of smoothness in all three dimensions (x, y, z). In our initial evaluations, we noted wide variability in the smoothness estimates in the z-direction, that was not related to our PVAF measurements, and so we report here only the geometric mean smoothness estimated in the x and y directions [mean = $\sqrt{(FWHMx*FWHMy)}$]. Smoothness estimation when the FWHM is small relative to the voxel size can be inaccurate due to discretization effects (Forman et al., 1995, Kiebel et al., 1999), but the method suggested by Forman et al. (1995) and implemented in AFNI is designed to provide accurate smoothness estimates even when FWHM is low. Smoothness estimates were based on brain voxels only. Non-brain voxels were masked prior to smoothness estimation. Smoothness was estimated prior to processing ("raw smoothness") and after processing (motion correction, time-shifting and spatial normalization) ("processed smoothness"). Unless specifically termed "processed smoothness", all smoothness estimates are raw smoothness.

Smoothness equalization—smoothing all scanners to a FWHM of 7 mm

After observing substantial differences in smoothness and related BOLD sensitivity between different sites, we hypothesized that smoothing all scanners to an equal level (see below) would reduce the differences in scanner BOLD effect sizes across scanners. A custom computer program was devised which iteratively assessed the smoothness of a data set and then smoothed it with a Gaussian kernel until the target smoothness was achieved. We chose to smooth all studies to a FWHM of 7 mm for reasons that are explained in the results section below. After informal testing, it was apparent that smoothness estimates provided by

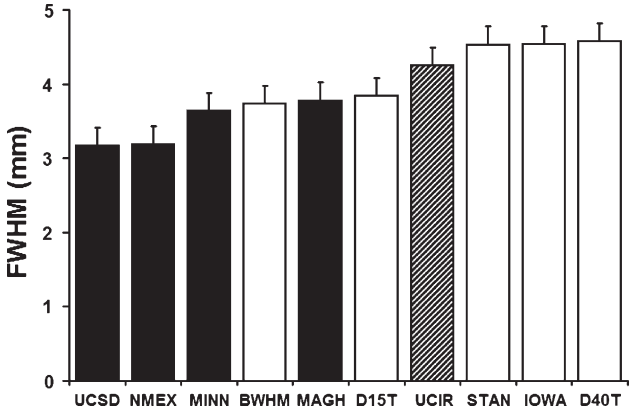

Fig. 2. Smoothness across scanners. The black bars indicate Siemens scanners, the open bars indicate GE scanners, and the slanted-line filled bar represents the single Picker scanner in the study. Smoothness is indicated as an estimated FWHM (in mm) of a Gaussian smoothing kernel. The error bar represents the standard error.

3dFWHM were remarkably stable across volumes within a run, and thus a single measurement (the 40th volume index of each run) would suffice during this iterative equalization step.

Evaluating the effect of acquisition type (Spiral vs. EPI) and K-space filtering on smoothness

To explain differences between scanners on smoothness, we conducted methodological tests on 2 subjects at the IOWA scanner using the same scanner used in the main study. This scanner was ideal, since it was one of the most smooth scanners in the study. The two factors that had been raised as possibly affecting smoothness were k-space filtering and acquisition type (Spiral vs. EPI). GE scanners employ a Fermi filter for k-space filtering (Lowe and Sorenson, 1997). For each subject, 4 runs were collected: (1) acquisition type=EPI, Fermi radius=32, Fermi width=10; (2) acquisition type=EPI, Fermi radius=32, Fermi width=2.5; (3) acquisition type=spiral-out, Fermi radius=32, Fermi width=2.5; (4) acquisition type=EPI, Fermi radius=256, Fermi width=1 (same as filter off). The acquisition parameters were identical to those of the main study. For each run, 50 TRs were collected, with no activation task, and were processed identically to images in the main study.

Statistical analysis

Data were available for 5 subjects at 10 scanners. Each subject was scanned on 2 visits, and there were 4 runs per visit. Visit and run effects were often negligible and were consequently dropped from certain analyses in the interest of model simplification. Almost all analyses were performed in SAS (Cary, NC) and most employed the PROC MIXED procedure (Littell et al., 1999), and Restricted Maximum Likelihood Estimate (ReML). Degrees of freedom for fixed effects were estimated using the calculations detailed by Kenward and Roger (1997). The scanners were divided into low-field strength scanners (five scanners at 1.5 T) and high-field strength scanners (four 3 T and one 4 T scanner).

Scanner effects on smoothness were assessed with scanner modeled as a fixed effect, whereas subject and subject-by-scanner were modeled as random effects. Field strength and vendor effects on smoothness were assessed with field strength (High vs. Low), vendor (GE vs. Siemens) and field-strength-by-vendor interactions modeled as fixed effects whereas subject, subject-by-field-strength and subject-by-vendor interactions were modeled as random effects. The relationship between smoothness and activation effect size (PVAF) was assessed with a random coefficients model. Scanner effects within a field strength were modeled with scanner as a fixed effect and subject, visit, subject-by-visit, subject-by-scanner, scanner-by-visit and scanner-by-subject-by-visit as random effects. Field strength effects were assessed with field (low vs. high) modeled as a fixed effect, whereas scanner, subject and visit and their interactions were modeled as random effects.

Results

Smoothness

Scanner effects

The average FWHM for each scanner is shown in Fig. 2. The scanner effect was highly significant (F(scanner)=111.0, df=9,36, $p<0.0001$). The smoothest scanner (D40T) was 1.44 times more smooth than the least smooth scanner (UCSD). This comparison contrasted a GE scanner running a spiral-out acquisition at 4.0 T to a Siemens 1.5 T running a conventional EPI sequence. Smoothness differences were somewhat greater among low-field scanners than among high-field scanners. The coefficient of variation (CV) of smoothness between low-field scanners was 0.16 whereas for high-field scanners the CV was 0.11. Furthermore, the ratio of smoothest-to-least-smooth among low-field scanners was 1.43 whereas the same ratio was only 1.26 for the high-field scanners (Fig. 2).

Field strength and manufacturer effects

High-field scanners (3.0 T and 4.0 T) were significantly smoother than low-field scanners (F=50.8, df=1,4.12, p=0.0018) (Fig. 3). The mean FWHM of the high-field scanners was 4.05 mm, the mean FWHM of the low-field scanners was 3.80 mm, and the ratio of high to low was 1.07. Furthermore, GE scanners were significantly more smooth than Siemens

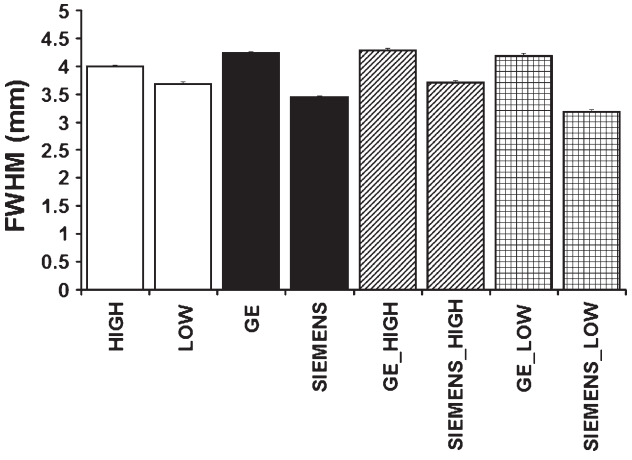

Fig. 3. Illustration of field strength and vendor effects on smoothness (FWHM, in mm). Least squares means (and standard errors) from the overall ANOVA analysis are plotted. The open bar comparison is for field strength, the dark-filled bar comparison is for vendor (GE vs. Siemens), the slanted-line bar comparison is for vendor-within-high-field and the cross-hatched bar comparison is for vendor-within-low-field.

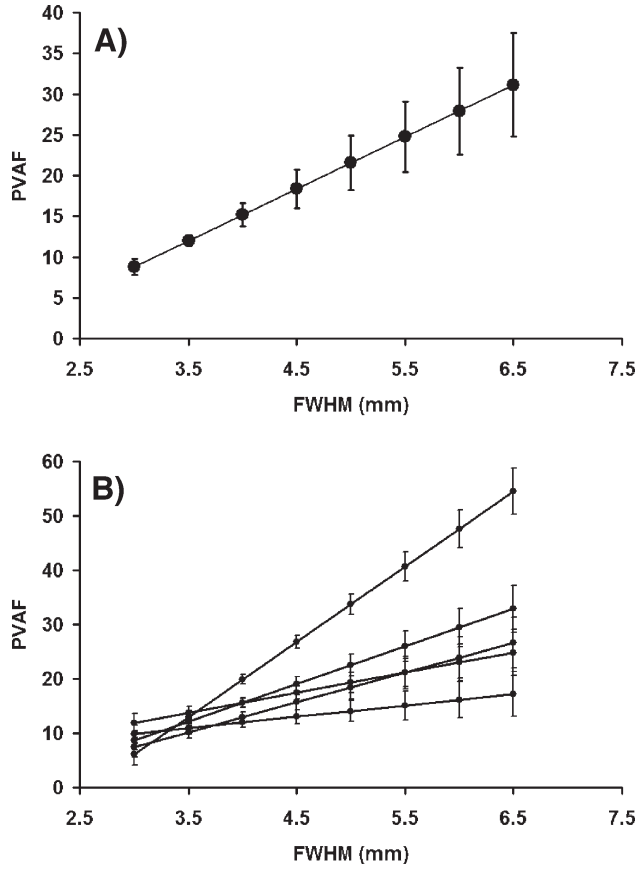

Fig. 4. (A) Plot of the relationship between raw smoothness (abscissa, FWHM, in mm) and PVAF (ordinate, activation effect size) across all 5 subjects, as estimated by the random coefficients regression of PVAF onto smoothness described in the text. (B) The same relationship as in panel A plotted for each subject individually. The error bars are standard errors of the estimates.

scanners ($F=525.8$, $df=1,348$, $p<0.0001$) (Fig. 3). The average FWHM of the GE scanners was 4.24 mm, the average FWHM of the Siemens scanners was 3.45 mm, and the ratio of FWHM between GE and Siemens scanners was 1.23. There was a field-strength-by-vendor interaction ($F=39.8$, $df=1,348$, $p<0.0001$). The significant interaction stems from the fact that the vendor effect is greater in low-field scanners than in high-field scanners (Fig. 3). The single Picker scanner was left out of this comparison.

Processed smoothness

The relationship between raw smoothness and processed smoothness was highly linear (Pearson $r=0.952$, $p<0.001$). When processed smoothness was modeled as the dependent measure and raw smoothness as the independent measure (intercept=0.0), the slope of the relationship was 1.25. The smoothest data were from the D40T scanner, subject 5, visit 1, run 4, with a processed FWHM of 6.83 mm. To test the effect of smoothness equalization on reducing variance due to scanner, we smoothed all the data to a FWHM of 7.0 mm, since this was just slightly higher than the largest FWHM estimate in the processed data.

The relationship between smoothness and activation effect size

Smoothness was significantly related to PVAF ($F=9.9$, $df=1,4$, $p=0.03$) (Fig. 4A). The slope of the relationship was 6.37, indicating that for every 1-mm increase in FWHM, one would expect a 6.4 increase in PVAF. The same relationship is plotted for each subject in Fig. 4B. Note that a positive relationship between FWHM and PVAF is noted for each subject.

Activation effect size (PVAF)

Factor analysis of the ROI data

In a effort to reduce the number of variables to be analyzed at this stage, PVAF data from all 8 ROIs (based on unsmoothed data) were entered into a factor analysis (SPSS, Inc). Only 1 factor (eigenvalue >1.0) was found, and all 8 ROIs were substantially

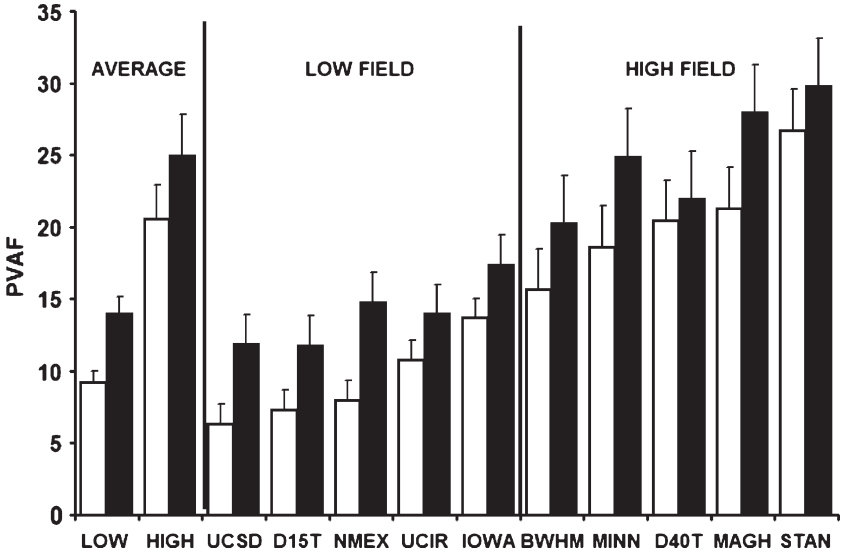

Fig. 5. PVAF results before and after smoothness equalization. The open bars represent data before smoothness equalization and the black bars represent data after smoothness equalization. In the left portion, estimated means for all low-field and all high-field scanners are shown. In the middle portion, data from each scanner in the low-field group are presented. On the right portion, data from each scanner in the high-field group are presented. The error bars are standard errors.

weighted on this factor. However, the initial communality for the bilateral visual cortex ROI was substantially lower (0.53) than that for the other 7 ROIs (all greater than 0.71). This may have been due to the fact that, in this initial study, display luminance and visual angle were somewhat dissimilar across scanners (data not shown). For these reasons, data from the visual cortex region were dropped from further analyses. For overall assessment of factors explaining variance in PVAF, the PVAF values for the remaining 7 ROIs were averaged. However, data on each of these 7 ROIs are also presented below.

Activation effect size (PVAF)

The mean activation for each scanner, before and after smoothness equalization, is shown in Fig. 5. Obviously, smoothness equalization resulted in increased activation effect size for all cases, as would be expected from Fig. 4. PVAF values for both low-field and high-field scanners increased after smoothness equalization (low: 4.7%, high: 4.4%). The difference in PVAF between low- and high-field scanners before and after smoothness equalization was similar (approximately 11%). However, the ratio PVAF of high-to-low-field scanners prior to smoothness equalization was 2.23 and after smoothness equalization was 1.79, a reduction of 20%. The scanner with the highest unsmoothed activation effect size was STAN (PVAF = 26.7%), and the scanner with the lowest unsmoothed activation effect size was UCSD (PVAF = 6.4%), an approximate 4-fold difference. Even within a field strength group, there was substantial variation from scanner to scanner in activation effect size. Prior to smoothness equalization, the ratio of activation effect size between the strongest low-field scanner (IOWA) and the weakest low-field scanner (UCSD) was 2.15. For high-field scanners, the ratio prior to smoothness equalization was 1.71. Smoothness equalization decreased these

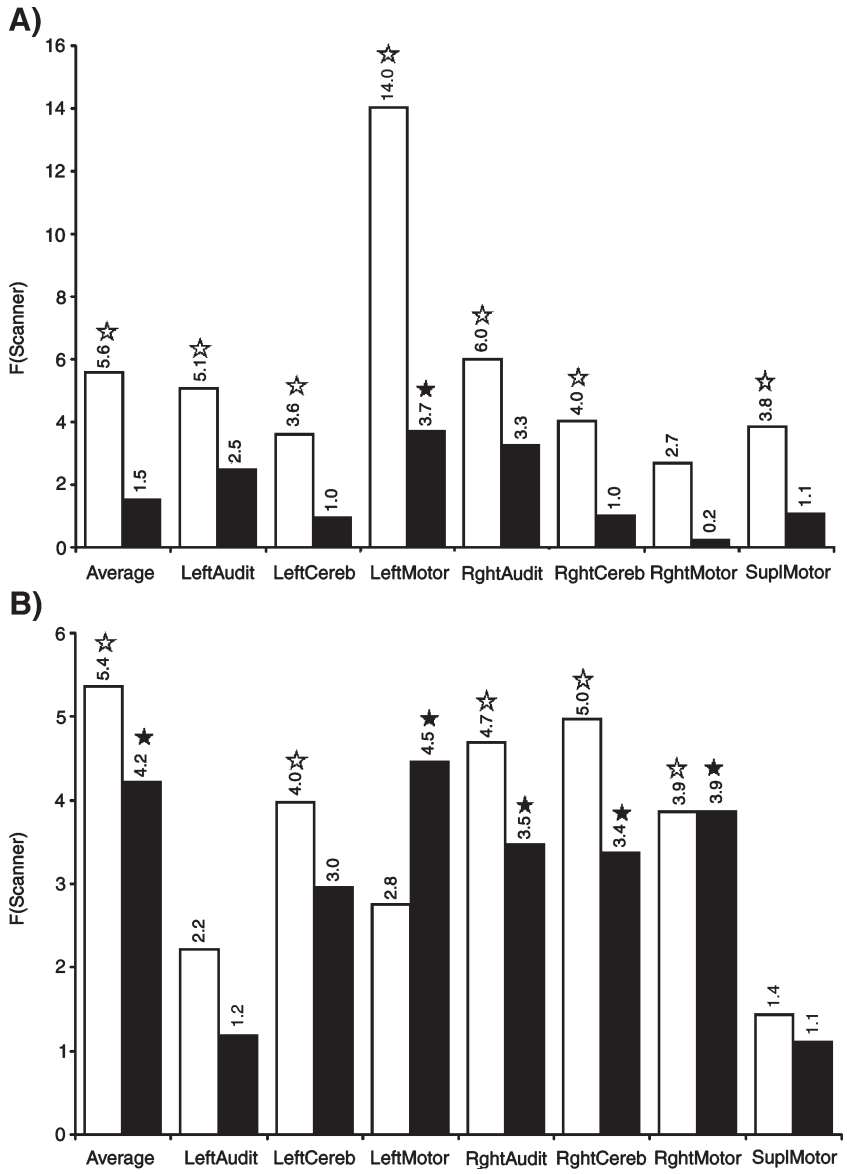

Fig. 6. (A) Plot of the *F* values for the scanner effect within the low-field scanners over the 8 measures (average and 7 ROIs). The open bars represent data before smoothness equalization and the dark bars represent data after smoothness equalization. Stars indicate statistical significance ($p < 0.05$). (B) Same as panel A but for high-field scanners only.

ratios by 32% in the low-field case and by 14% in the high-field case. Note also that standard error estimates are higher among high-field scanners, as would be expected since PVAF measures are also higher (Fig. 5).

F values for scanner effects

The *F* values for the scanner effect within field strength group are shown in Figs. 6A and B. Among low-field scanners, smoothness equalization reduced the scanner effect *F* value for the average measure as well as for each of the 7 remaining ROIs. Furthermore, among the low-field scanners, the *F*-tests for the scanner effect were statistically significant for all regions in the unsmoothed state (8 of 8 comparisons), whereas after smoothness equalization, the *F*-tests were significant only for the left motor ROI (Figs. 6A and B).

Among high-field scanners, smoothness equalization reduced the scanner effect *F* value for the average measure as well as for 5 of the 7 remaining ROIs. Among the high-field scanners, the *F*-tests for the scanner effect were statistically significant for 5 of 8 measures in the unsmoothed state and after smoothness equalization.

Coefficient of variation between scanners

To further illustrate the effect of smoothness equalization on reducing variation among scanners, we computed coefficients of variation (CV) for low-field scanners (Fig. 7A) and for high-field scanners (Fig. 7B). Data for the average measure and for each of the 7 ROIs are presented. For low-field scanners, the CV decreased in every measure by an average of 47%. For high-field scanners, the CV decreased in 7 of 8 measures by an average of 13%. Thus,

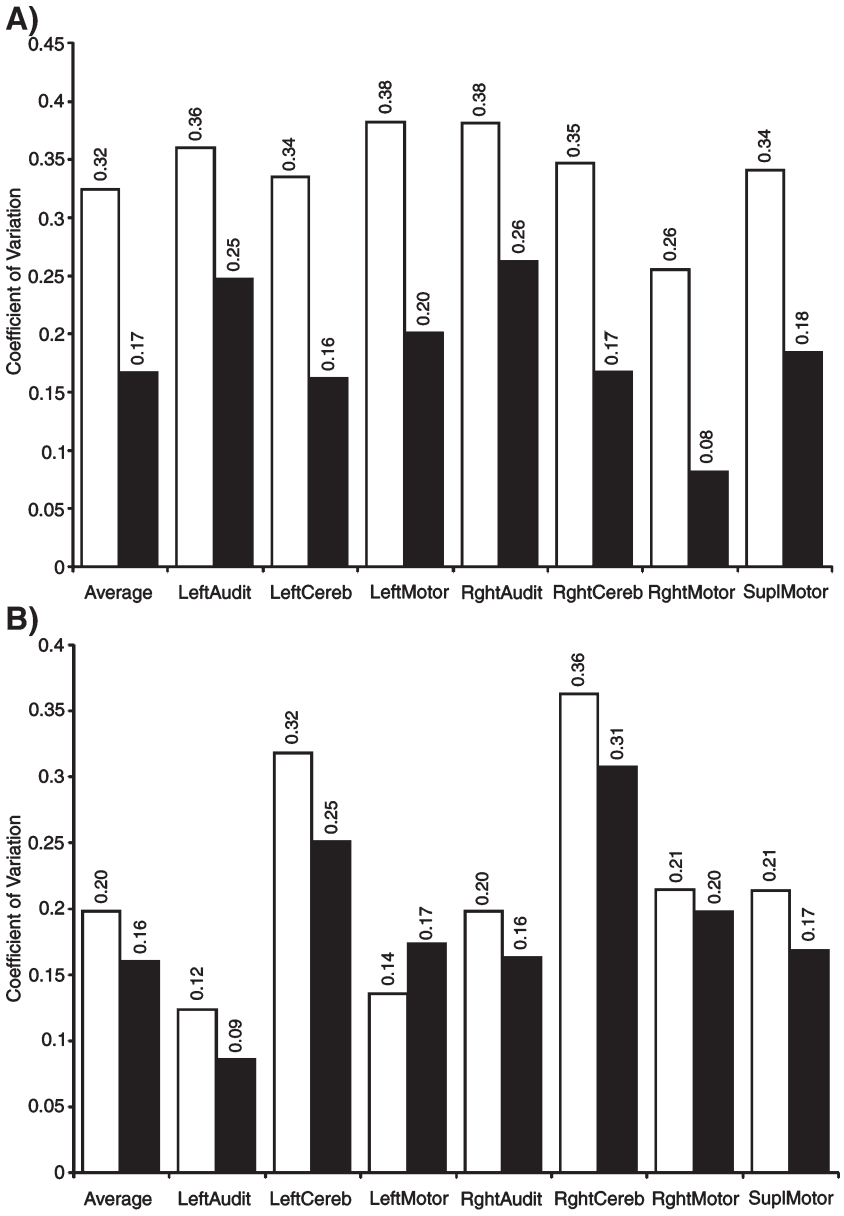

Fig. 7. (A) Plot of the coefficient of variation (CV) values among the low-field scanners over the 8 measures (average and 7 ROIs). The open bars represent data before smoothness equalization and the dark bars represent data after smoothness equalization. (B) Same as panel A but for high-field scanners only.

the effect of smoothness equalization on reducing scanner-to-scanner variation is about 3.5-fold greater among low-field scanners than among high-field scanners (Figs. 7A and B).

Field strength effects

The *F* values for the field strength effect are presented in Fig. 8 for the eight measures. It is apparent that smoothness equalization decreased the magnitude of the field strength effect *F* value on activation effect size for all measures. Nonetheless, the field strength effect was statistically significant after smoothness equalization for all ROIs except for the supplementary motor cortex. Thus, smoothness equalization does not eliminate the strong field strength effects (Fig. 8).

Evaluating the effect of acquisition type (spiral vs. EPI) and K-space filtering on smoothness

Plots for several k-space filter functions are presented in Fig. 9. K-space filtering appeared to have the dominant effect on spatial smoothness (Table 2). Results for both subjects followed the same pattern: the least smooth images were reconstructed with essentially no k-space filtering (Fermi radius=256. Fermi width=1). EPI images reconstructed with a Fermi radius of 32 and a Fermi width of 2.5 were substantially smoother than the unfiltered EPI images and were smoother than spiral-out images reconstructed with the same filter settings. The smoothest functional images resulted from EPI reconstructed with a Fermi width=32 and a Fermi radius=10. The smoothness, vendor, acquisition type and filter settings for each scanner in the study are listed in Table 3, sorted by decreasing smoothness.

Discussion

This large multicenter study, in which the same five healthy normal subjects traveled to 9 sites and were evaluated with fMRI on 10 scanners, has provided an unprecedented opportunity to evaluate the influence of scanner differences on fMRI results. We report marked, highly statistically significant scanner differences in smoothness and activation effect size. We also show that the smoothness of the images produced by different scanners does have an important relationship to activation effect size. Results demonstrate that smoothness equalization, i.e., smoothing the image data from all scanners to a constant FWHM (here 7 mm), reduces scanner differences in activation—an effect which is most marked in low-field strength scanners.

One key difference between scanners, within and between field strength groups, was the smoothness of the raw images produced. The smoothest scanner had a FWHM 1.44 times greater than that of the least smooth scanner, and the differences between scanners on smoothness were highly statistically significant. We also report that functional images from GE scanners were significantly smoother than functional images from Siemens scanners. To the best of our knowledge, this is the first report of this vendor effect.

Our small study of the effect of k-space filtering and acquisition-type (EPI vs. Spiral) clearly suggests that differences in scanner smoothness are strongly a function of k-space filtering (Tables 2 and 3). Most probably, the key reason why the GE scanners are smoother than Siemens scanners has to do with the different k-space filtering algorithms employed. GE functional sequences (both EPI and spiral), by default, employ a rotationally symmetric k-space filter (Fermi filter, Lowe and Sorenson, 1997). For the standard EPI sequence on a GE scanner, assuming a 64×64 matrix, the default Fermi filter width is 32 and the default radius is 10. The spiral sequence written by GHG and supplied to GE users in the present study uses a width of 2.5, which is a more suitable filter width. The Fermi filter design is especially desirable because it can be very flat in the passband. With a reasonable radius (32) and width setting (2.5), the filter passes most of spatial frequencies of interest, but attenuates the corners of k-space. This is necessary if one is to have a rotationally symmetric point-spread function for each voxel. However, with a width of 10, as used in the stock GE EPI sequence employed by the IOWA scanner, desirable spatial frequencies not in the corner of k-space are also attenuated (Fig. 9). This will impart substantial unnecessary smoothness to the final, reconstructed image. Siemens, on the other hand, does not employ any k-space

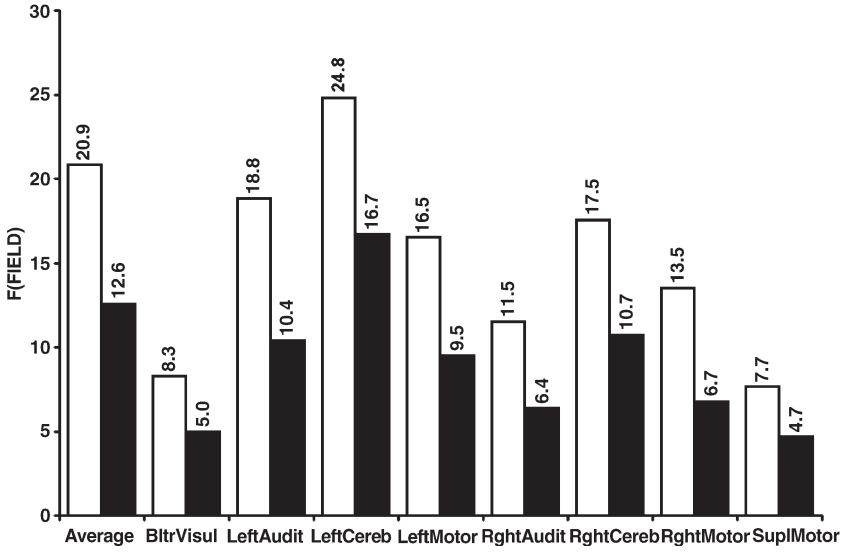

Fig. 8. Plot of the *F* values for the field strength effect over the 8 measures (average and 7 ROIs). The open bars represent data before smoothness equalization and the dark bars represent data after smoothness equalization. The only *F* value that was not statistically significant ($p<0.05$) was for the supplementary motor cortex after smoothness equalization.

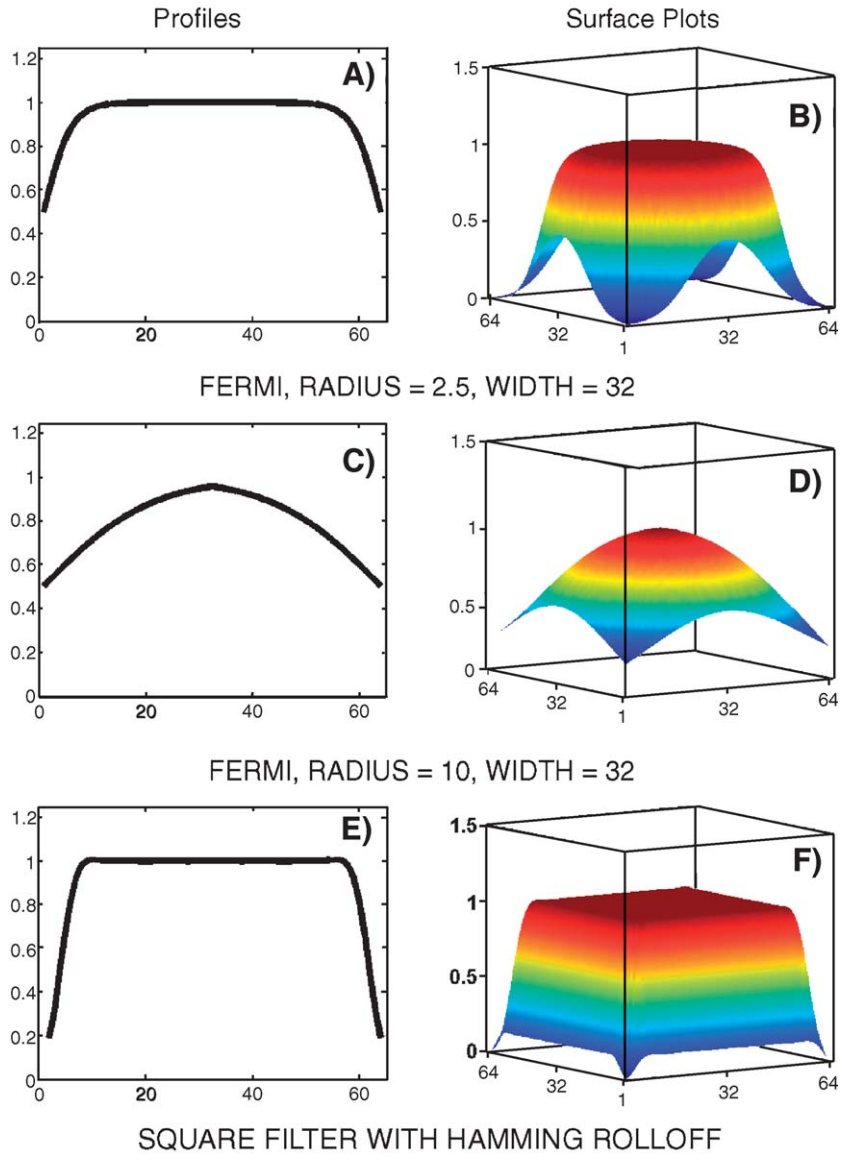

Fig. 9. Profiles and surface plots for 3 key k-space filters applied in this study. Profile (A) and surface plot (B) of a Fermi filter with a radius of 32, and a width of 2.5. Profile (C) and surface plot (D) of a Fermi filter with a radius of 32, and a width of 10. Profile (E) and surface plot (F) of a custom filter applied to the data from the single scanner employing a Pickar scanner (UCIR).

filtering on its EPI sequences by default. Although a k-space filter can be applied, none was employed by any Siemens site in the present study. Although this does not result in a rotationally symmetric point-spread function, it does result in an overall increase in spatial resolution of the resulting images. We also found that functional images collected with a spiral-out acquisition were not smoother than images collected with an EPI acquisition when the Fermi filter settings were identical.

Table 2
Study of K-space filtering and acquisition type

Acquisition type	Fermi_Radius	Fermi_Width	Average FWHM	
			Subject 1	Subject 2
EPI	32	10	4.19	4.00
EPI	32	2.5	4.01	3.89
Spiral-Out	32	2.5	3.82	3.74
EPI	256	1	3.05	2.96

We found that high-field scanners in our study are 1.07 times smoother that low-field scanners. Yang et al. (1999) compared effective pixel dimensions at 4 T to 1.5 T, based on data from two GE scanners running spiral acquisitions and reported a ratio of 1.2. When we make the same 4T/1.5 T comparison, using the 2 Duke/UNC scanners (GE-spiral), we find a ratio of 1.18. Krasnow et al. (2003) compared smoothness estimates from a GE 3 T to a GE 1.5 T scanner, both running spiral acquisitions, and found a ratio of 1.14. In the Krasnow et al. (2003) study, the smoothness difference was also reported to be statistically significant (p values range from 0.0095 to 0.00002). The increased smoothness of high-field scanners is completely expected, since T2 decay is faster at 3 T than 1.5 T, and thus there is typically more signal decay during the

Table 3
Smoothness, vendor, acquisition type and K-space filter settings for the 10 scanners—sorted by decreasing smoothness (FWHM)

Site	Average FWHM	Vendor	Sequence type	K-space filter	Fermi filter settings	
					Radius	Width
D40T	4.58	GE	Spiral-Out	Yes	32	2.5
IOWA	4.54	GE	EPI	Yes	32	10
STAN	4.53	GE	Spiral-In/Out	Yes	32	2.5
UCIR	4.25	Picker	EPI	Hamming[a]		
D15T	3.84	GE	Spiral-Out	Yes	32	2.5
MAGH	3.78	Siemens	EPI	No		
BWHM	3.74	GE	EPI	No		
MINN	3.64	Siemens	EPI	No		
NMEX	3.19	Siemens	EPI	No		
UCSD	3.18	Siemens	EPI	No		

[a] The k-space filtering employed at this scanner was not rotationally symmetric, but rectangular (Fig. 9). It was flat in the passband, with a very sharp rolloff at the edges of k-space.

readout phase (Farzaneh et al., 1990). Such decay has the effect of decreasing spatial resolution at high-field strengths. Removing smoothness differences between high- and low-field scanners by smoothing all studies to a FWHM of 7 mm (smoothness equalization) resulted in a 20% reduction of the activation effect size ratio of high to and low-field scanners, but high-field scanners were still statistically significantly more sensitive than low-field scanners.

As noted in the Introduction, we thought it important to assess smoothness in these scanners since there is a well known relationship between spatial smoothness and increased SNR (Parrish et al., 2000). The present report further documents a highly significant relationship between scanner smoothness and activation effect size. A smoothness increase of 1 mm (FWHM) corresponded to an increase of approximately 6.4% increase in activation effect size measured as PVAF. It is not uncommon for fMRI researchers to smooth by 8 mm FWHM or more, and our data suggest that this level of post-acquisition smoothing should produce very substantial increases in activation effect size.

Even within a field strength group (low: 1.5 T, high: 3T/4 T), there were statistically significant activation effect size differences between scanners in the unsmoothed state. In both field strength groups, the strongest-to-weakest ratios of activation effect size were in the 2.0 range. The *F*-tests for the scanner effects on activation effect size were statistically significant in 7 of 8 measures for low-field scanners and 5 of 8 measures for high-field scanners. Thus, not all 1.5 T scanners are alike and not all high-field scanners are alike.

Smoothness equalization did indeed reduce scanner effects on activation effect size, but this result was strongest for low-field scanners. The strongest-to-weakest ratios within a field strength group were reduced by 32% in low-field strength scanners and 14% in high-field strength scanners. The *F* values for the scanner effect were reduced for all 8 measures among low-field scanners and for 6 of 8 measures for high-field scanners. After smoothness equalization, the *F* values for the scanner effect were statistically significant for only a single region (left motor ROI) among low-field scanners, whereas for high-field scanners, the *F*-tests for the scanner effect was still significant for 5 of 8 measures. The coefficient of variation between scanners within a field strength group was decreased by 47% among low-field scanner and by 13% among high-field strength scanners. Thus, smoothness equalization is an excellent tool to reduce scanner-induced variance in activation effect size, particularly among low-field scanners. The reduced effectiveness of smoothness equalization among high-field scanners is probably partially related to the somewhat reduced smoothness variation among this scanner group.

It is entirely unsurprising that high-field scanners are more sensitive than low-field scanners—this is consistent with theoretical considerations (Kruger et al., 2001) and much empirical data (Turner et al., 1993; Bandettini et al., 1994; Gati et al., 1997; Yang et al., 1999; Krasnow et al., 2003; Fera et al., 2004). In terms of the current analysis, high-field scanners were 2.23 times more sensitive than low-field scanners prior to smoothness equalization and 1.79 times more sensitive after smoothness equalization. These ratios are comparable to 3T/1.5 T ratios reported by Kruger et al. (2001) for CNR (1.8 to 2.2) and mean *z*-score (1.7 to 1.9) within activated areas. Although exact numerical comparisons with published reports is not completely valid, given different study characteristics, this field strength effect appears to be much larger than that reported by Yang et al. (1999) at 4 T and by Fera et al. (2004) at 3 T, both of which, by our calculations[2] appear to report effect size ratios of 1.2 or less. It is not clear why our results and the results of Kruger et al. (2001) indicate a much stronger field strength effect than the studies by Yang et al. (1999) and Fera et al. (2004).

Although smoothness equalization markedly reduced activation effect size differences between scanners, others factors are undoubtedly contributing to PVAF, especially among the high-field scanners. First, the high-field strength group was not as homogeneous as the low-field strength group, since the high-field strength group included a 4 T scanner. Also, MAGH employed a double-echo EPI sequence and STAN employed a spiral in/out sequence (Glover and Law, 2001). These double acquisitions are averaged in some manner during final image reconstruction and this averaging should increase activation effect size[3]. Indeed these 2 scanners had the strongest activation effect size. Head coil performance can obviously affect activation effect size and at least 1 scanner (STAN) employed a custom built head coil designed to enhance SNR.

It can be argued that smoothness equalization across multiple sites results in a lower resolution group scan characteristic. However, the majority of fMRI analyses are already being performed with a smoothing step during preprocessing, because of the assumptions of Gaussian Random Field theory needed for some algorithms, or simply to increase statistical power at the expense of resolution. Thus, the suggested smoothing equalization does not necessarily result in a significant degradation in net resolution for the study. With our adaptive algorithm, smoother scanners require less filtering than less smooth scanners, providing

[2] Fera et al. (2004) report an optimum *t* value for finger tapping activation of motor cortex of $t=3.4$ at 1.5 T and $t=4.0$ at 3 T. With 82 TRs ($df=80$) these translate into effect sizes of 0.76 and 0.89 respectively, for a ratio of 3 T/1.5 T of 1.18. Yang et al. (1999) report an optimum *t* value for finger tapping activation of motor cortex of $t=5.77$ at 4 T and 4.82 at 1.5 T. Given 84 TRs at 4 T and 48 TRs at 1.5 T, this translates into effect sizes of 1.27 for 4 T and 1.42 at 1.5 T, yielding a ratio of 0.9—the low field scanner actually has larger effects!

[3] The second acquisition of these paired acquisition sequences should be smoother since it is acquired later, when signal decay can lead to blurring. This could lead to increased smoothness in the final images.

a more homogeneous resolution data set for group analyses. Of course, if there are outlier scanners in the group, it would be desirable to take steps to improve the resolution of those sites a priori, e.g., by adjusting local k-space filters.

Post hoc smoothness equalization is not the only strategy to deal with scanner differences in smoothness. Given our findings on the importance of k-space filtering, it would certainly be advantageous for all scanners to use exactly the same filter and filter settings. Another approach would be to tailor nominal voxel sizes at each scanner so that each scanner produced an effective voxel size that is equivalent. One disadvantage of post hoc smoothness equalization is that it requires that all scanners be smoothed to that of the most smooth scanner in the set. Since smoothing is tantamount to throwing away the outer regions of k-space data, it results in a lower scan efficiency and an SNR disadvantage relative to a strategy where lower resolution k-space data are acquired but for the same readout duration (i.e., at reduced net bandwidth). Thus, using the "voxel-size adjustment" approach, one could set smoother scanners to collect a smaller nominal voxel size by adjusting the field of view (within limits). With some trial and error, the effective voxel size of the smooth scanners could be matched to the effective voxel size of the unsmooth scanners.

In summary, we have described important scanner differences in activation effect size and smoothness that will affect the results of multicenter fMRI studies. Vendor differences in image smoothness were documented and are likely due to differences in k-space filtering regimes. We have shown that smoothness equalization can reduce scanner differences in activation effect size within a field strength and also reduce the field strength effect on activation effect size.

Acknowledgments

This research was supported by a grant [#5 MOI RR 000827] to the FIRST Biomedical Informatics Research Network (BIRN, http://www.nbirn.net), that is funded by the National Center for Research Resources (NCRR) at the National Institutes of Health (NIH).

The members of the FIRST BIRN project all deserve acknowledgement for their significant efforts, but unfortunately, they are too numerous to mention. Please visit http://www.nbirn.net for more information regarding key personnel. We would identify two people for special acknowledgement: Dr. Jessica Turner, the project coordinator, worked tirelessly to facilitate the entire project; and the FIRST BIRN principal investigator, Dr. Steven Potkin, who provided vision and leadership throughout. We would also like to thank Drs. Gunnar Krüger and Stefan Thesen (both from Siemens AG, Erlangen, Germany) for help with the interpretation of some of the MR physics underlying our findings.

References

Bandettini, P., Wong, E., Jesmanowicz, A., Prost, R., Cox, W., Hinks, R., Hyde, J., 1994. MRI of human brain activation at 0.5 T, 1.5 T, and 3.0 T: comparisons of R2* and functional contrast to noise ratio [oral]. 2'nd Proc. Soc. Magn. Reson., San Francisco, p. 434.

Berry, I., Barker, G.J., Barkhof, F., Campi, A., Dousset, V., Franconi, J.M., Gass, A., Schreiber, W., Miller, D.H., Tofts, P.S., 1999. A multi-center measurement of magnetization transfer ratio in normal white matter. J. Magn. Reson. Imaging 9, 441–446.

Casey, B.J., Cohen, J.D., O'Craven, K., Davidson, R.J., Irwin, W., Nelson, C.A., Noll, D.C., Xiaoping, H., Lowe, M.J., Rosen, B.R., Truwitt, C.L., Turski, P.A., 1998. Reproducibility of fMRI results across four institutions using a spatial memory task. NeuroImage 8, 249–261.

Cohen, M.S., 1997. Parametric analysis of fMRI data using linear systems methods. NeuroImage 6, 93–103.

Cox, R., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.

de Certaines, J.D., Henriksen, O., Spisni, A., Cortsen, M., Ring, P.B., 1993. In vivo measurements of proton relaxation times in human brain, liver, and skeletal muscle: a multicenter MRI study. Magn. Reson. Imaging 11, 841–850.

Farzaneh, F., Riederer, S.J., Pelc, N.J., 1990. Analysis of T2 limitations and off-resonance effects on spatial resolution and artifacts in echo-planar imaging. Magn. Reson. Med. 14 (1), 123–139.

Fera, F., Yongbi, M.N., van Gelderen, P., Frank, J.A., Venkata, M.S., Duyn, J.H., 2004. EPI-BOLD fMRI of human motor cortex at 1.5 T and 3.0 T: sesitivity dependence on echo time and acquisition bandwidth. J. Magn. Reson. Imaging 19, 19–26.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C., 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn. Reson. Med. 33 (5), 636–647.

Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., Turner, R., 1998. Event-related FMRI: characterizing differential responses. NeuroImage 7 (1), 30–40.

Gati, J.S., Menon, R.S., Ugurbil, K., Rutt, B.K., 1997. Experimental determination of the BOLD field strength dependence in vessels and tissue. Magn. Reson. Med. 38 (2), 296–302.

Glover, G.H., Law, C.S., 2001. Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. Magn. Reson. Med. 46, 515–522.

Kenward, M.G., Roger, J.H., 1997. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53 (3), 983–997.

Kiebel, S.L., Poline, J.-B., Fiston, K.J., Holmes, A.P., Worsley, K.J., 1999. Robust estimation in statistical parametric maps using standardized residuals from the general linear model. NeuroImage 10, 756–766.

Krasnow, B., Tamm, L., Greucius, M.D., Yang, T.T., Glover, G.H., Reiss, A.L., 2003. Comparison of fMRI activation at 3 and 1.5 T during perceptual cognitive, and affective processing. NeuroImage 18, 813–826.

Kruger, G., Kastrup, A., Glover, G.H., 2001. Neuroimaging at 1.5 T and 3.0 T: comparison of oxygenation-sensitive magnetic resonance imaging. Mag. Reson. Med. 45, 595–604.

LaConte, S., Anderson, J., Muley, S., Ashe, J., Frutiger, S., Rehm, K., Hansen, L.K., Yacoub, E., Hu, X., Rottenburg, D., Strother, S., 2003. The evaluation of preprocessing choices in single-subject BOLD fMRI using NPAIRS performance metrics. NeuroImage 18, 10–27.

Liao, C., Worsley, K.J., Poline, J.-B., Aston, J.A.D., Duncan, G.H., Evans, A.C., 2002. Estimating the delay of the fMRI response. NeuroImage 16, 593–606.

Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., 1999. SAS Systems for Mixed Models. SAS Institute Inc, North Carolina.

Lowe, M.J., Sorenson, J.A., 1997. Spatially filtering functional magnetic resonance imaging data. Megn. Reson. Med. 37, 723–729.

Paley, M., Cozzone, P.J., Alonso, J., Vion-Dury, J., Confort-Gouny, S., Wilkins, I.D., Chong, W.K., Hall-Craggs, M.A., Harrison, M.J., Gili, J., Rovira, A., Capellades, J., Rio, J., Ocana, I., Nicoli, F., Dhiver, C., Gastaut, J.L., Gastaut, J.A., Wicklow, K., Sauter, R., 1996. A multicenter proton magnetic resonance spectroscopy study of neurological complications of AIDS. AIDS Res. Hum. Retroviruses 12, 213–222.

Parrish, T.B., Gitelman, D.R., LaBar, K.S., Mesulann, M.M., 2000. Impact of signal-to-noise on functional MRI. Magn. Reson. Med. 44 (6), 925–932.

Schnack, H.G., Van Haren, N.E.M., Hulshoff Pol, H.E., Picchioni, M., Wiesbrod, M., Saurer, H., Cannon, T., Huttenun, M., Murray, R., Kahn, R.S., 2004. Reliability of brain volumes from multi-center MRI acquisition: a calibration study. Hum. Brain Mapp. 22, 313–320.

Silver, N.C., Barker, G.J., Miller, D.H., 1999. Standardization of magnetization transfer imaging for multi-center studies. Neurology 53 (5), S33–S39.

Strother, S., La Conte, S., Hansen, L.K., Anderson, J., Zhang, J., Pulapura, S., Rottenburg, D., 2004. Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. A preliminary group analysis. NeuroImage 23, S196–S207.

Turner, R., Jezzard, P., Wen, H., Kwong, K.K., Le Bihan, D., Zeffiro, T., Balaban, R.S., 1993. Functional mapping of the human visual cortex at 4 and 1.5 tesla using deoxygenation contrast EPI. Magn. Reson. Med. 29, 277–279.

Van Haren, N.E.M., Cahn, W., Hulshoff Pol, H.E., Schnack, H.G., Caspers, E., Lemstra, A., Sitskoorn, M.M., Wiersma, D., van den Bosch, R.J., Dingemans, P.M., Schene, A.H., Kahn, R.S., 2003. Brain volumes as predictor of outcome in recent-onset schizophrenia: a multi-center MRI study. Schizophr. Res. 64, 41–52.

Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G.H., Morales, F., Evans, A.C., 2002. A general statistical analysis for fMRI data. NeuroImage 15, 1–15.

Yang, Y., Wen, H., Mattay, V.S., Balaban, R.S., Frank, J.A., Duyn, J.H., 1999. Comparison of 3D BOLD functional MRI with spiral acquisition at 1.5 T and 4.0 T. NeuroImage 9, 446–451.

Zou, K.H., Greve, D.N., Wang, M., Pieper, S.D., Warfield, S.K., White, N.S., Manandhar, S., Brown, G.G., Vangel, M.G., Kikinis, R., Wells III, W.M., 2005. Reproducibility of functional MR imaging: preliminary results of prospective multi-institutional study performed by biomedical informatics research network. Radiology 237, 781–789.