AJNR

Reproducibility of Visual Activation in Functional MR Imaging and Effects of Postprocessing

Atsushi Miki, Jonathan Raz, Theo G. M. van Erp, Chia-Shang J. Liu, John C. Haselgrove and Grant T. Liu

Reproducibility of Visual Activation in Functional MR Imaging and Effects of Postprocessing

Atsushi Miki, Jonathan Raz, Theo G. M. van Erp, Chia-Shang J. Liu, John C. Haselgrove, and Grant T. Liu

BACKGROUND AND PURPOSE: **Functional MR imaging studies of the brain should be interpreted in the context of their reproducibility. We assessed the reproducibility of visual activation measured by functional MR imaging and analyzed the effect of image transformation to standard space.**

METHODS: **Seven healthy volunteers were studied twice with echo-planner functional MR imaging at 1.5 T during visual stimulation. The studies were separated by an interval of 2 to 7 days. Functional images were analyzed after spatial normalization to the space described by Talairach and Tournoux and/or after coregistration of the images of the second study with the images of the first study. The number of active voxels for each study was determined at three thresholds. In addition, the change in the center of the mass of activation, the mean change in signal intensity, and the mean *t* value within the activated area were measured. These reproducibility indexes were calculated for the spatially normalized and nonnormalized data for each subject.**

RESULTS: **Variations in visual activation were observed between the two studies in the same individual as well as across subjects. There was no evidence of an effect from image transformation on reproducibility on any of the measures.**

CONCLUSION: **Our findings show that the reproducibility of activation in functional MR imaging may be much more variable across subjects than suggested in previous studies. The use of different types of image transformation (coregistration, spatial normalization) does not significantly affect the reproducibility of visual activation.**

Functional MR imaging is a noninvasive and relatively accessible tool by which to investigate human brain function with high spatial and temporal resolution (1, 2). Its noninvasiveness and availability permit repetitive scans in the same subject. However, the reliability of functional MR imaging as measured by test/retest reproducibility and its

ability to detect subtle changes in a subject's condition (eg, visual function) have not been established conclusively. Since robust visual activation by functional MR imaging can be observed and the retinotopic organization in the visual cortex has been demonstrated in detail by functional MR imaging (3), the visual cortex seems to be the ideal brain region in which to test the reproducibility of functional MR imaging. To evaluate the test/retest reproducibility of visual activation, we prospectively studied multislice echo-planar functional MR images in healthy volunteers who were scanned on two different days by using the same visual activation paradigm in a 1.5-T scanner. We measured and compared the location and the magnitude of the visual activation at three thresholds for each study. In addition, we analyzed the functional images after spatial normalization and/or after coregistration of the images of the second study with the images of the first study and investigated the effects of these procedures on reproducibility of several measures.

Methods

Subjects and Data Acquisition

Seven healthy volunteers (four men and three women, 22–27 years old; mean age, 24 years) gave informed consent

before participating in this study. The consent form was approved by the institutional review board of the Children's Hospital of Philadelphia. All subjects had normal visual acuity, confrontational visual fields, and stereopsis. No subject had a history of visual loss or neurologic disease. All subjects were examined twice in the same manner and underwent the second session 2 to 7 days (mean, 4.5 days) later. All studies were performed between 6 >/SCAP<AM>/SCAP< and 8 >/SCAP<AM>/SCAP<.

Imaging was performed with a clinical 1.5-T MR system. The magnet was shimmed using an automatic shimming routine with first- and second-order gradients. The subjects' heads were cushioned with foam padding within the quadrature head coil to restrict motion. Subjects were instructed to hold their heads still. Identical midsagittal images were acquired using the slice acquisition procedure developed by Noll and colleagues (4) to reproduce the anatomic and functional images across sessions. First, coronal scout images were obtained, and oblique axial images perpendicular to the midline of the previous coronal images were acquired to account for head tilt. Subsequently, sagittal images perpendicular to the midline of the previous oblique axial images were acquired to account for head rotation. Finally, 16 oblique axial images positioned parallel to the calcarine fissure were obtained to encompass the visual cortex. The relative angles between the anterior commissure-posterior commissure line and the selected planes were recorded to ensure that the images in the first and second studies were acquired with the identical orientation. The angle ranged from 14° to 23° (mean, 19°). The lowest slice was positioned at the anterior commissure. The anatomic images were obtained using a T1-weighted spin-echo sequence with the following parameters: TR/TE = 500/15, matrix = 256 × 256, field of view = 240 mm, in-plane resolution = 0.94 × 0.94 mm², slice thickness = 3 mm, gap = 2 mm. Thereafter, 16 functional images were acquired with slices identical and parallel to those of the anatomic images by using a T2*-weighted echo-planar imaging sequence (TR/TE = 1.68/64, flip angle = 90°, matrix = 64 × 64, field of view = 240 mm, in-plane resolution = 3.75 × 3.75 mm², slice thickness = 5 mm, no interslice gap). In all, 120 sets of 16 images each were acquired for functional imaging at interscan intervals of 3 seconds. The acquisition period for the functional images consisted of 12 epochs. Light-proof binocular goggles with 5 × 6 light-emitting diodes (modified S10VSB, Grass Instruments, Quincy, MA) flashing at a frequency of 8 Hz were placed over the subjects' eyes to provide binocular full-field visual stimulation. The subjects were instructed to keep their eyes open during the period of visual stimulation. The visual stimuli were turned on and off with the use of a trigger from the magnet. Ten scans obtained during visual stimulation of both eyes (epochs 1, 3, 5, 7, 9, and 11) alternated with 10 scans obtained during darkness (epochs 2, 4, 6, 8, 10, and 12).

Data Analysis

Data analysis was performed on UNIX workstations. IDL and SPM96 (Wellcome Department of Cognitive Neurology, London, UK) packages were used. The first five scans of the echo-planar images were discarded to eliminate magnetic saturation effects. The average signal intensity of each image in the functional imaging set was normalized to compensate for baseline drift of the MR signal. Functional images of each subject were realigned using a six-parameter (three translations and three rotations) rigid body transformation. To test whether realignment strategies affected test/retest reproducibility, the images from the second study were realigned to the first volume of the first study or to that of the second study. Furthermore, the effect of spatial normalization was examined by processing the images with and without transformation into the anatomic space described by Talairach and Tournoux (5). This spatial normalization routine was performed by minimizing the sum of squared difference between the functional images and the echo-planar imaging template, using an eight-parameter affine transformation. Data were smoothed with a gaussian filter

(full width at half maximum = 8.0 × 8.0 × 10.0 mm). A boxcar function delayed by 6 seconds and temporal smoothing was used, and t statistics were calculated for each voxel and then transformed into Z values (SPM[Z]). The Z map was thresholded with a Z value of 3.5, 4.5, or 5.5 to define activated areas. In all subjects, a Z value greater than 4.5 approximately corresponded to $P < .05$ after correction for multiple comparisons in the entire image. The number of supra-threshold voxels was tabulated, and the center of mass of the supra-threshold voxels was determined at these three thresholds. Average signal intensity changes and average Z values were obtained for the supra-threshold (Z > 3.5, 4.5, or 5.5) voxels.

Statistical Analysis of Reproducibility

To evaluate reproducibility, R_{size} (the ratio of active volumes [min/average]) (6) and $R_{overlap}$ (the ratio of the common area to the average) (6) were calculated using the following equations:

$$R_{size} = 2 \times \frac{V_{min}}{(V_1 + V_2)}$$

$$R_{overlap} = 2 \times \frac{V_{overlap}}{(V_1 + V_2)}$$

where V_{min} is the smaller of the V_1 (activated volume in the first study) and V_2 (activated volume in the second study). $V_{overlap}$ was the volume activated by both the first and second studies of each subject (6). Difference indexes, defined as the squared difference of two values from each study divided by the sum of the values, were calculated for the mean signal intensity changes and mean Z scores within the activated areas, respectively. Additionally, the distance between the center of mass on the Z maps (Z > 3.5, 4.5, or 5.5) of the two studies was calculated. Visual activation of the nonspatially normalized data of the first study was compared with the activation of the second study coregistered to the first study (type 1 comparison). The spatially normalized data of the first study were compared with the spatially normalized data of the second study without coregistration to the first study (type 2 comparison), and the spatially normalized data of the first study and spatially normalized data of the second study with coregistration to the first study (type 3 comparison) were also compared. An approximate F-test in a linear mixed effects model was performed to evaluate the effects of the thresholds and the type of comparisons on the reproducibility indexes of visual activation. Additionally, the number of voxels above the threshold and the mean Z values within the area of activation for each threshold in each study were compared.

Results

The area of activation during visual stimulation was confined mainly to the primary visual cortex (Brodmann's area 17), especially on the medial side of the occipital lobe in all seven subjects. The center of the mass of activation lay almost on the midline, which suggested that stimulation of both eyes produced symmetrical activation of the bilateral visual cortex in most subjects. The location of the activation suggests that these areas represent true visual activation. However, the volume of activation varied considerably among subjects and within the same subject at different times (Figs 1 and 2). The activated volumes ranged from 10 to 1220 voxels for the threshold of Z > 4.5. Detailed results are available from the authors upon request.
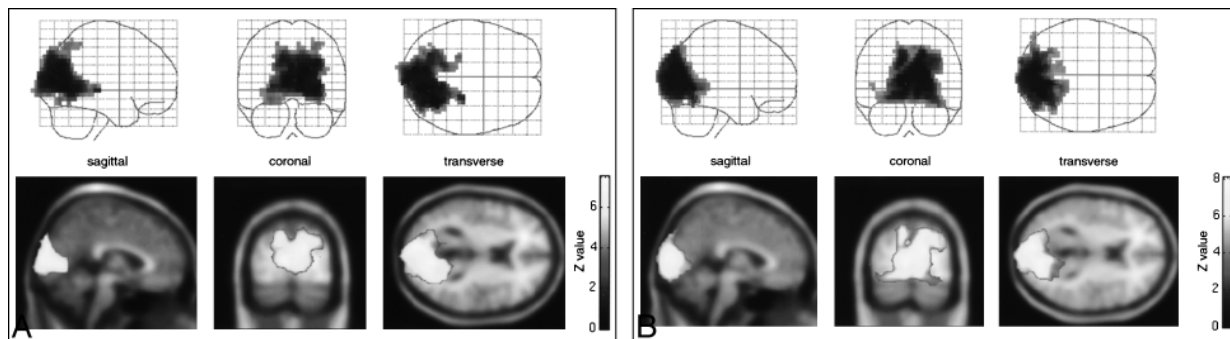
FIG 1.   A and B, Z maps of the first (A) and second (B) studies in a volunteer show high reproducibility (type 2 comparison). A height threshold of uncorrected P < .001 was selected for this figure. Supra-threshold clusters of voxels were identified using a corrected spatial extent threshold of P < .05.
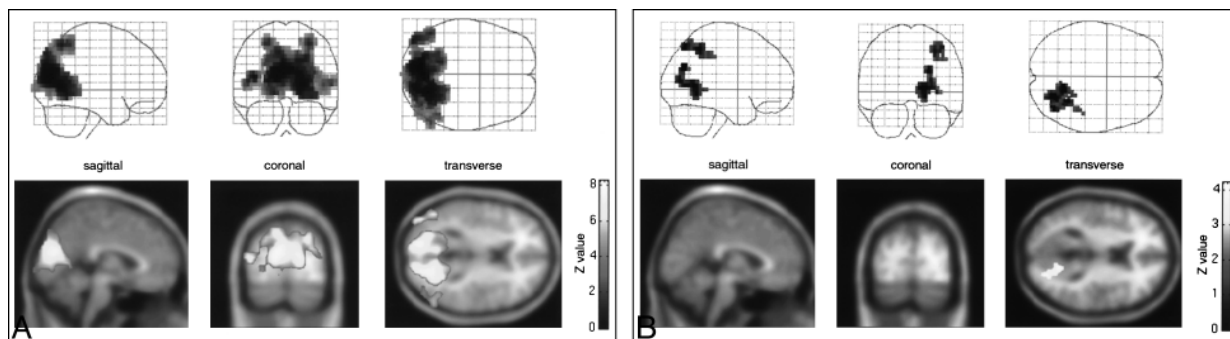

FIG 2.   A and B, Z maps of the first (A) and second (B) studies in a volunteer show lower reproducibility (type 2 comparison). This subject had much less visual activation in the second study.

TABLE 1: The ratio of the size of the supra-threshold areas (R_{size})

Type of Comparison	Threshold	Mean	Standard Deviation	Minimum	Maximum
1	Z > 3.5	0.67	0.33	0.19	0.95
1	Z > 4.5	0.60	0.40	0.04	0.95
1	Z > 5.5	0.59	0.43	0.01	0.97
2	Z > 3.5	0.68	0.34	0.12	0.97
2	Z > 4.5	0.61	0.39	0.02	0.94
2	Z > 5.5	0.58	0.42	0.00	0.99
3	Z > 3.5	0.68	0.35	0.18	0.97
3	Z > 4.5	0.61	0.40	0.03	0.96
3	Z > 5.5	0.59	0.43	0.00	1.00

TABLE 2: The ratio of the common area between the supra-threshold voxels for the first and second studies ($R_{overlap}$)

Type of Comparison	Threshold	Mean	Standard Deviation	Minimum	Maximum
1	Z > 3.5	0.48	0.29	0.10	0.76
1	Z > 4.5	0.48	0.33	0.02	0.77
1	Z > 5.5	0.44	0.35	0.00	0.76
2	Z > 3.5	0.50	0.26	0.09	0.74
2	Z > 4.5	0.49	0.30	0.01	0.75
2	Z > 5.5	0.45	0.32	0.00	0.74
3	Z > 3.5	0.49	0.27	0.12	0.73
3	Z > 4.5	0.47	0.32	0.02	0.73
3	Z > 5.5	0.43	0.34	0.00	0.73

Summaries of the statistics computed from R_{size} and $R_{overlap}$ appear in Tables 1 and 2, respectively. Both these indexes have large standard deviations and wide ranges, suggesting prominent intersubject variability in the reproducibility of visual activation as measured by functional MR imaging.

The F-test revealed weak evidence that R_{size} (P = .0654) and $R_{overlap}$ (P = .0671) were affected by thresholds. R_{size} and $R_{overlap}$ were greatest at the lowest threshold (Z > 3.5) (Figs 3 and 4). There was weak evidence for the effect of the type of comparison on $R_{overlap}$ (P = .069). Type 2 comparison (spatial normalization without coregistra-

tion) had the greatest $R_{overlap}$ (Fig 4). There was also weak evidence for an interaction (P = .0598) between the type of comparison and thresholds on the difference index for the signal intensity changes. There were no significant effects of the thresholds and the type of comparisons on other reproducibility indexes, except for an effect of the thresholds on the mean Z values. There was no significant study effect on the number of activated voxels and mean Z scores of the activated voxels for any of the three types of comparisons.
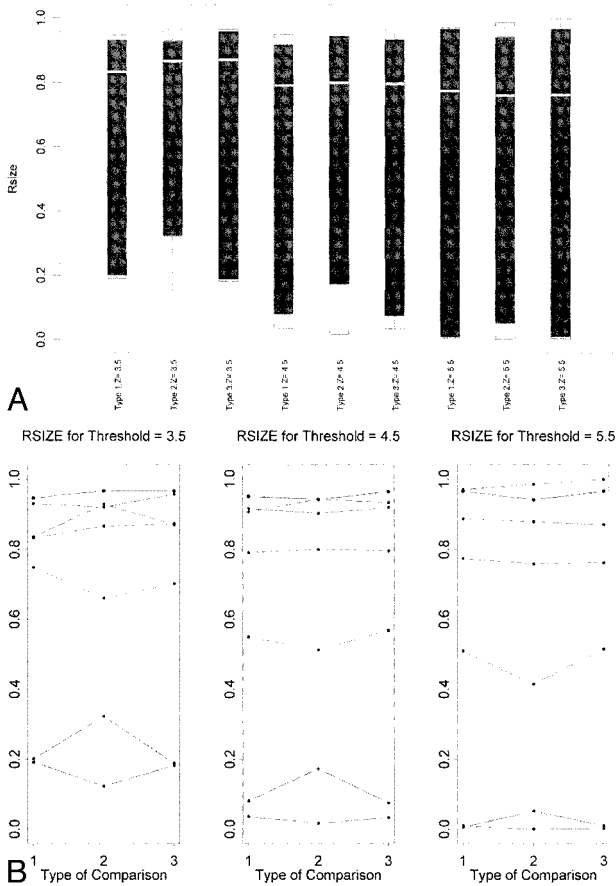
FIG 3. *A* and *B,* Box plots of R_{size} for seven subjects (*A*) and plots for each subject (*B*) show the effects of the type of comparison and threshold value on reproducibility. *Black bars* in *A* represent the R_{size} values of the upper quartile (75th percentile) to the lower quartile (25th percentile). The median (50th percentile) values for each threshold are represented by the *white lines* in the bars. The minimum and maximum values are connected to the black bar with *whiskers* (the *dotted lines* extending from the bar). The *brackets* at both ends represent the maximum and minimum values. Where the maximum or minimum value exceeds the upper or lower quartile, respectively, by 1.5 times the difference of the values of the upper and lower quartiles, the extreme point is shown separately from the bar without the whiskers. The *points* in *B* represent the R_{size} values of each individual in each type of comparison, with *dotted lines* connecting each individual's set of values across the types of comparison.

FIG 4. *A* and *B,* Box plots of $R_{overlap}$ for seven subjects (*A*) and plots for each subject (*B*).

Discussion

Intersubject variability has been reported in many functional MR imaging studies. Such variability may reflect not only attention or alertness of the subjects during the paradigm but also anatomic variation among individuals (7). The tremendous structural variation prevents us from analyzing functional MR imaging data quantitatively across subjects without using spatial transformation to a common standard stereotactic space, which reduces spatial resolution. Accordingly, intrasubject comparisons are thought to be more valid than intersubject comparisons in functional MR imaging. However, our analysis shows that visual activation in functional MR imaging can be variable even within the same subject.
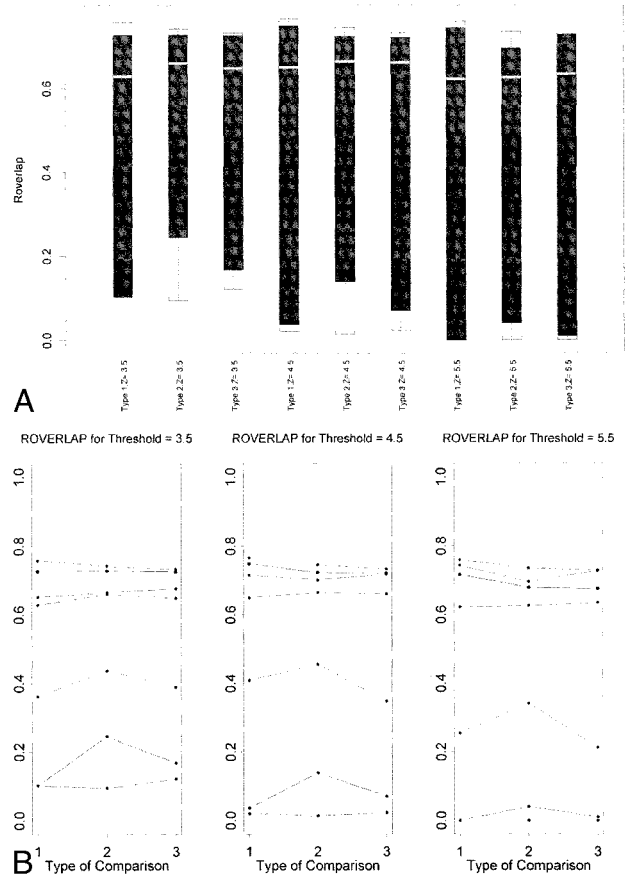
Test/retest reproducibility (or reliability) within the same subject has been evaluated by several investigators within a session (4, 6, 8–13) and across sessions (4, 6, 8, 10, 12, 14–16). Two measurements can be performed without moving the subject in and out of the scanner; however, it is impossible to obtain two data sets on separate days without repositioning the subject, which may affect reproducibility. We attempted to minimize repositioning errors by employing the algorithm of Noll et al (4). We conducted this study to determine reproducibility across sessions, which parallels the clinical application in which patients are examined on two separate occasions to assess changes in brain activation during longitudinal follow-up. Because subjects were exposed to the same paradigm in the second session as they were in the first, some adaptation may have occurred between the first and second procedures, resulting in smaller activated areas in the second experiment. However, we did not find evidence of an exposure effect on the number and mean Z scores of the voxels above thresholds.

Although visual inspection is a subjective way to evaluate reproducibility, several investigators (8, 12, 14) have used this method to report the reproducibility of activation in functional MR imaging. They found good correspondence in respect to lo-

cation and size of activated areas among data sets. We found very good agreement in some subjects but also noticed large variations of activated areas in other subjects. However, since this method of data analysis is not quantitative, it is difficult to draw any reliable conclusions as to the reproducibility of functional MR imaging by using visual inspection alone. Therefore, quantitative analysis is desirable.

Functional MR imaging provides us with information not only about the extent and location of the activated area but also about the magnitude of the response to the task. In this study we analyzed reproducibility on the basis of the ratio of the active volumes, the ratio of the common area to the average of the two studies, and displacement of the center of the mass of activation. These parameters revealed spatial information about activation. The mean Z values and mean signal intensity changes disclosed the magnitude of the response. Ramsey et al (10) showed that the magnitude of significant signal change was consistent across trials, whereas the number of activated voxels was variable. Moser et al (11) also reported better reproducibility for signal enhancement than for the number of activated pixels.

Test/retest reproducibility may be influenced by several factors: differences in the subjects' condition and position, instability of the MR scanner, errors in data processing and statistics, the type of stimulus used to elicit a response, and the attention level of the subjects. These effects are usually confounded and difficult to separate from one another. Although repositioning errors should have been minimized in this study by the careful matching procedure (4), it might have been difficult to obtain exactly the same volume twice. Several investigators (4, 6, 10) have compared the reproducibility of intersession data with that of intrasession data, and all but one (10) found better reproducibility for the intrasession data. This may be due to a repositioning error or to a change in the condition of the subject or machine. Rombouts et al (6) found that the reproducibility of visual cortex activation increased with the use of a gaussian filter and an increase in filter width.

Healthy volunteers were used in this study. Because patients, especially those with neurologic defects, may not be as cooperative as healthy people, the reproducibility found here may not apply to patients. Also, reproducibility may be different among the tasks used for activation. Activation by cognitive tasks may be less robust, requiring the acquisition of more images to achieve a signal-to-noise ratio similar to that found with activation by simple sensory or motor tasks (4). The visual stimulus we used in this study was a simple flash, but a more complex stimulus might either enhance or decrease reproducibility of visual activation. Additional studies are necessary to evaluate reproducibility of functional MR imaging using different kinds of visual stimulation. Since the subject's at-

tention by itself is known to alter activation in the primary visual cortex (17), additional efforts to control attention may be necessary to increase reproducibility.

Usually, statistical thresholding at a particular value is performed to identify regions of activation in functional MR imaging experiments. However, there is a trade-off between sensitivity and specificity when using this method. Thresholding at a low value may include truly nonactive areas and thresholding at a high value may miss truly active areas (10). The variation in reproducibility could be due to the particular measures we used, since they are dependent on thresholds. Yetkin et al (9) used correlation thresholds of 0.50, 0.60, and 0.70, and found higher reproducibility at the lower threshold than at the higher threshold. Rombouts et al (6) varied the significance level (range, 0.20–0.95) and found that maximum reproducibility was obtained with a significance level just below a *P* value of .05, after Bonferroni correction. Although our results show only weak evidence of the effects of the thresholds on reproducibility, this may be due to our relatively small sample size. Nevertheless, our results are in good agreement with those of Rombouts et al (6), in that the threshold Z value of 3.5 had the greatest reproducibility and the Z value of 4.5 corresponded approximately to $P = .05$ after correction for multiple comparisons. Moser et al (11) demonstrated that the application of an adaptive threshold resulted in better reproducibility than did a fixed threshold. Noll et al (4) suggested a method for determining an optimal threshold by using receiver operator characteristic curves, but they required multiple (more than three) trials per subject to estimate several parameters.

Although it is difficult to compare our reproducibility indexes with those of other investigations, owing to differences in the data analysis procedures, task paradigms, and scanning parameters, the study by Rombouts et al (6) is similar to ours in some aspects. The functional images were taken with echo-planar imaging sequences in two different sessions, and visual stimulation was used in both studies. However, our reproducibility values are worse than those reported by Rombouts and colleagues. The average R_{size} was 0.88 and 0.70 for $R_{overlap}$ (full width at half maximum, 8 mm) in their study, but those values were 0.60 and 0.48 in our study (Z > 4.5). In particular, the test/retest reproducibility of two of the subjects in our sample was much worse than that in the previous report (6). In one subject, R_{size} ranged from 0.01 to 0.32, and in the other from 0.00 to 0.19. These values are far lower than those in the previous report, and cannot be explained by a small sample size. Therefore, our study indicates that the variability in reproducibility across subjects may have been underestimated. Another study, which investigated functional MR imaging during a finger opposition task, showed an average R_{size} of 0.38 and an average $R_{overlap}$ of 0.31 (10), which seems to suggest that reproducibility of

sensorimotor activation may be less than that of visual activation in functional MR imaging.

Spatial normalization is a useful method for reporting significantly activated areas by their locations after normalizing them into the same space, usually the one described by Talairach and Tournoux (5). This technique has been used for reporting the location of activation in the lateral geniculate nucleus (18) and the V4 area (19), and is increasingly used in analyses of functional MR imaging studies. To our knowledge, no report has examined the reproducibility of functional MR imaging data after spatial normalization. The difference in the reproducibility between nonnormalized and normalized data was small; that is, spatial normalization did not substantially change reproducibility.

All subjects except for one showed good activation in the whole primary visual cortex, with activation extending from the posterior visual cortex to the anterior visual cortex. Thus, although good agreement in visual activation between the two studies was not obtained in some subjects, most subjects seem to have had true activation in at least one of the sessions, as expected from people with normal visual function. Therefore, in some instances, it may be necessary to repeat the acquisition of functional MR images twice to confirm the results. In one subject, the activated area was confined to the posterior visual cortex in both studies, which was not in accordance with the subject's normal visual fields. Although the reason for this occurrence is unknown, it may be that this subject is among the nonresponders in functional MR imaging studies (10, 15).

Conclusion

Reproducibility of visual activation in functional MR imaging varies across studies, even in the same subject. Therefore, care should be taken when interpreting the results of functional MR imaging studies, even when the same subject is being investigated repeatedly. We did not find evidence of an effect on reproducibility from image transformation, such as in spatial normalization to the standard brain or in coregistration of one image to another.

References

1. Kwong KK, Belliveau JW, Chesler DA, et al. **Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation.** *Proc Natl Acad Sci U S A* 1992;89:5675–5679
2. Ogawa S, Tank DW, Menon R, et al. **Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging.** *Proc Natl Acad Sci U S A* 1992;89:5951–5955
3. Sereno MI, Dale AM, Reppas JB, et al. **Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging.** *Science* 1995;268:889–893
4. Noll DC, Genovese CR, Nystrom LE, et al. **Estimating test-retest reliability in functional MR imaging: 2, application to motor and cognitive activation studies.** *Magn Reson Med* 1997;38:508–517
5. Talairach J, Tournoux P. *Co-planar Stereotaxic Atlas of the Human Brain.* Stuttgart: Thieme; 1988
6. Rombouts SARB, Barkhof F, Hoogenraad FGC, et al. **Within-subject reproducibility of visual activation patterns with functional magnetic resonance imaging using multislice echo planar imaging.** *Magn Reson Imaging* 1998;16:105–113
7. Stensaas SS, Eddington DK, Dobelle WH. **The topography and variability of the primary visual cortex in man.** *J Neurosurg* 1974;40:747–755
8. Binder JR, Rao SM, Hammeke TA, et al. **Localized human brain language systems demonstrated by task subtraction functional magnetic resonance imaging.** *Arch Neurol* 1995;52:593–601
9. Yetkin FZ, McAuliffe TL, Cox R, Haughton VM. **Test-retest precision of functional MR in sensory and motor task activation.** *AJNR Am J Neuroradiol* 1996;17:95–98
10. Ramsey NF, Tallent K, Van Gelderen P, Frank JA, Moonen CTW, Weinberger DR. **Reproducibility of human 3D fMRI brain maps acquired during a motor task.** *Hum Brain Mapping* 1996;4:113–121
11. Moser E, Teichtmeister C, Diemling M. **Reproducibility and postprocessing of gradient-echo functional MRI to improve localization of brain activity in the human visual cortex.** *Magn Reson Imaging* 1996;14:567–579
12. Huckins SC, Turner CW, Doherty KA, Fonte MM, Szeverenyi NM. **Functional magnetic resonance imaging measures of blood flow patterns in the human auditory cortex in response to sound.** *J Speech Lang Hear Res* 1998;41:538–548
13. Lee CC, Jack CR, Riederer SJ. **Mapping of the central sulcus with functional MR: active versus passive activation tasks.** *AJNR Am J Neuroradiol* 1998;19:847–852
14. DeYoe EA, Bandettini P, Neitz J, Miller D, Winans P. **Functional magnetic resonance imaging (FMRI) of the human brain.** *J Neruosci Methods* 1994;54:171–187
15. Rombouts SARB, Barkhof F, Hoogenraad FGC, Sprenger M, Valk J, Scheltens P. **Test-retest analysis with functional MR of the activated area in the human visual cortex.** *AJNR Am J Neuroradiol* 1997;18:1317–1322
16. Wexler BE, Fulbright RK, Lacadie CM, et al. **An fMRI study of the human cortical motor system response to increasing functional demands.** *Magn Reson Imaging* 1997;15:385–396
17. Watanabe T, Sasaki Y, Miyauchi S, et al. **Attention-regulated activity in human primary visual cortex.** *J Neurophysiol* 1998;79:2218–2221
18. Buchel C, Turner R, Friston K. **Lateral geniculate activations can be detected using intersubject averaging and fMRI.** *Magn Reson Med* 1997;38:691–694
19. McKeefry DJ, Zeki S. **The position and topography of the human colour centre as revealed by functional magnetic resonance imaging.** *Brain* 1997;120:2229–2242