AJNR Am J Neuroradiol 24:652–657, April 2003

Reproducibility of Functional MR Imaging Results Using Two Different MR Systems

Erik-Jan Vlieger, Cristina Lavini, Charles B. Majoie, and Gerard J. den Heeten

BACKGROUND AND PURPOSE: **In the application of functional MR imaging for presurgical planning, high reproducibility is required. We investigated whether the reproducibility of functional MR imaging results in healthy volunteers depended on the MR system used.**

METHODS: **Visual functional MR imaging reproducibility experiments were performed with 12 subjects, by using two comparable 1.5-T MR systems from different manufacturers. Each session consisted of two runs, and each subject underwent three sessions, two on one system and one on the other. Reproducibility measures D (distance in millimeters) and R_{size} and $R_{overlap}$ (ratios) were calculated under three conditions: same session, which compared runs from one session; intersession, which compared runs from different sessions but from the same system; and intermachine, which compared runs from the two different systems. The data were averaged per condition and per system, and were compared.**

RESULTS: **The average same-session values of the reproducibility measures did not differ significantly between the two systems. The average intersession values did not differ significantly as to the volume of activation (R_{size}), but did differ significantly as to the location of this volume (D and $R_{overlap}$). The average intermachine reproducibility did not differ significantly from the average intersession reproducibility of the MR system with the worst reproducibility.**

CONCLUSION: **The location of activated voxels from visual functional MR imaging experiments varied more between sessions on one MR system than on other MR system. The amount of the activated voxels is independent of the MR system used. We suggest that sites performing functional MR imaging for presurgical planning measure the intersession reproducibility to determine an accurate surgical safety margin.**

In the past few years, interest in applying functional MR imaging in a clinical setting has grown, in particular in neuro-oncology where functional MR imaging could be used as a helpful tool for presurgical planning (1–4). However, this application would require that the functional MR imaging results be highly reproducible. Nowadays, most hospitals are equipped with more than one MR system, possibly even from different manufacturers. This raised the question as to whether the reproducibility of functional MR imaging results is independent of the MR system that was used to acquire the functional MR imaging data. Another area in which this could be important is in multicenter functional MR imaging studies.

The issue of reproducibility has been extensively addressed (5–18), but to the best of our knowledge the reproducibility resulting from the subsequent use of different MR systems has not been investigated so far. As our institution is equipped with two comparable 1.5-T MR systems from different manufacturers, we decided to repeatedly run the same functional MR imaging experiment with the same subjects by using the two different systems. Reproducibility experiments performed with visual stimuli usually yield higher reproducibility than do experiments with motor or language stimuli. Although language mapping and motor mapping are more important in the field of neurosurgery, we decided to use visual stimuli, as this would allow for a better distinction between what was inherent variability and what was added by the use of different systems. In this study, we wanted to quantify the added variability (if any) of using different but comparable MR systems on the reproducibility of functional MR imaging.

Received March 26, 2002; accepted after revision November 8.

From the Department of Radiology, Academic Medical Center, Amsterdam, the Netherlands.

Address reprint requests to Erik-Jan Vlieger, MD, Academic Medical Center, Department of Radiology, G1–209, P.O. Box 22660, 1100 DD, Amsterdam, the Netherlands.

Methods

Subjects

This study was performed with approval of the institutional medical ethics committee. Twelve healthy volunteers (three women, nine men; mean age, 27 years; range, 22–48 years)

participated in this study. Informed consent was obtained from all participants.

Experimental Protocol

Each subject underwent three sessions on three different days, twice on one system and once on the other. A session consisted of two runs, without head repositioning, so that each subject underwent six runs in total. In every session, a high-resolution 3D T1-weighted data set was acquired between the two runs. Five subjects had two sessions on system A and one on system B, and seven subjects had two sessions on system B and one on system A.

Experimental Setup

Visual stimuli were presented to the subjects by using the Integrated Functional Imaging System (MR imaging Devices Corporation, Waukesha, WI). This system consisted of a liquid crystal display (LCD) mounted above the head coil, connected by optical fibers to a computer placed outside the magnet room. The stimuli were generated by Eprime (version 1.0β5; Psychology Software Tools, Pittsburgh, PA). The size of the LCD screen was 17 × 13 cm, and the screen was positioned at an effective distance from the eyes of 40 cm. This yielded a viewing angle of 22°. Mean luminance was 7.8 cd/m^2 and contrast was 82%. If necessary, ocular refraction was corrected with MR-compatible glasses.

The block-design paradigm consisted of blocks of 45 seconds. The experimental condition consisted of an 8-Hz radial flickering black-and-white checkerboard, which was alternated with a black screen (the rest condition). In both conditions, a small white cross was presented at the center of the screen, and subjects were instructed to focus on this cross. Both the experimental and the rest conditions were repeated twice.

Section Positioning

Full-brain coverage was achieved by using a stack of 22 axial sections, without angulation. In functional MR imaging studies, more advanced section-positioning procedures have been described (9, 19), but these procedures may last up to 15 minutes per subject. We considered this time to be unacceptably long in a clinical situation and decided not to use those methods. Because of our section-positioning protocol, we expected an additional reproducibility error due to the partial-volume effect.

MR Imaging Acquisition

Two 1.5-T MR imaging systems were used: Magnetom Vision (software Numaris VB33; Siemens, Erlangen, Germany) and Signa Horizon (software LX 8.3; GE Medical Systems, Milwaukee, WI). Both systems had echo-planar imaging capabilities. On both systems, the standard quadrature head coils were used. Single-shot echo-planar imaging was used for functional imaging, with the following parameters for both systems: 4500/66 (TR/TE), 90° flip angle, 230-mm field of view, number of sections = 22, 128 × 128 matrix (full k-space), frequency direction left to right, 5-mm section thickness, no intersection gap, 189-second imaging time, bandwidth ≈ 1 kHz/pixel. For anatomic reference, a 3D T1-weighted image was acquired. On the Siemens system, the magnetization-prepared rapid acquisition gradient-echo (MP-RAGE) sequence was used (9.7/4/300 [TR/TE/TI], 12° flip angle), and on the GE system the 3D fast spoiled gradient-recalled acquisition in the steady state (FSPGR) sequence was used (30/6, 45° flip angle). In the remainder of the article, the brands of the systems are unnamed.

Image Processing and Statistical Analysis

The steps described in this section were performed for each subject separately. Images were transferred to a Pentium PC, and the BrainVoyager software (version 4.4; Brain Innovation, Maastricht, the Netherlands [20]) was used for image registration, motion correction, smoothing, and generation of statistical maps. Each echo-planar imaging time-course series (ETC) was registered to the anatomic volume from the same session by using section-position information, after which the anisotropic ETCs were interpolated to a 1-mm^3 resolution.

The six functional ETCs from the three separate sessions were registered to each other. As the first two ETCs were from the same session and no head repositioning was performed, they were already registered to each other. The first step to register the ETCs from the second session to the ETCs from the first session was to determine the rigid body transformation that aligned the second anatomic image to the first. This was accomplished by using the BrainVoyager software. The second step was to apply this transformation to the ETCs from the second session. The ETCs from third session were registered in the same way.

The ETCs were corrected for motion, and then spatial smoothing (full width at half maximum = 4 mm^3) was applied. Statistical maps (z-scores) were calculated by using the Student's t test. Activated voxels were defined as those voxels above a z-score threshold of 4 and a cluster-size threshold of 250 mm^3 (21). We restricted the analysis to the part of the brain posterior to the corpus callosum (13).

From the statistical maps, the following properties were calculated: the volume of the activated set of voxels, the average signal intensity change of these voxels (between rest and activated state), and the applied motion correction, both as to translation and as to rotation. These quantities were averaged per machine; we tested whether or not these averages differed significantly between the two systems by using the Mann-Whitney Test with SPSS software for Windows version 11.0 (SPSS Inc., Chicago, IL).

In the literature, different measures of reproducibility have been proposed (5–11, 13, 15). We chose to use the measures introduced by Rombouts et al (7). These measures are the ones most often used and therefore lend themselves best for comparison with previous work. In short, the measures used are D, R_{size}, and $R_{overlap}$, and they are used to compare two maps with activated voxels. With the center of an activated area defined as the center of mass of the set of activated voxels, D is the distance (in millimeters) between the centers of the activated areas, R_{size} is the ratio of the volumes of activated areas, and $R_{overlap}$ is the ratio of the area activated in both runs and the sum of the areas activated in each run separately: $R_{size} = 2[V_{min}/(V_1 + V_2)]$ and $R_{overlap} = 2[V_{overlap}/(V_1 + V_2)]$, where V_{min} is the smaller of V_1 and V_2 (the activated volumes of the first and the second study), and $V_{overlap}$ is the volume activated in both studies. Both R_{size} and $R_{overlap}$ range from 0 (completely unreproducible) to 1 (completely reproducible). D, R_{size}, and $R_{overlap}$ were calculated with homemade software.

Statistical maps were compared only within subject, and three different compare conditions were used: 1) maps from the same session (same session), 2) maps from different sessions but from the same system (intersession), and 3) maps from the two different systems (intermachine). For each system, comparisons were made between same-session and intersession reproducibility figures (of all the subjects), the same-session results were compared between the two systems as were the intersession results, and the intermachine results were compared with the intersession results of each system separately. Comparisons between the reproducibility values were made by using the Mann-Whitney Test, and P values were calculated. Errors reported in this article are standard deviations. Statistical significance was obtained when $P < .05$.

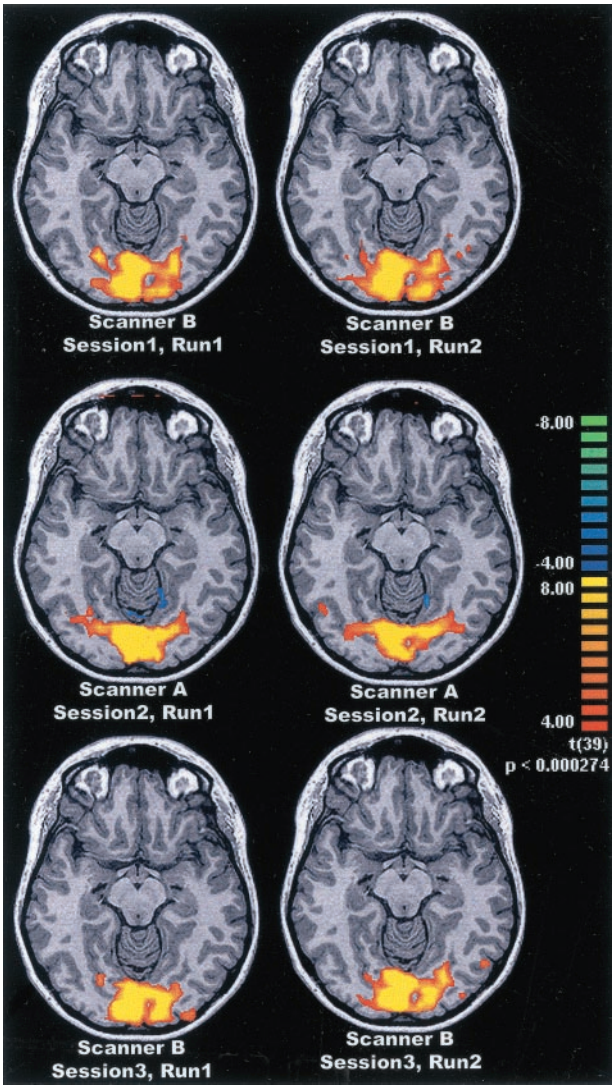

FIG 1. Example of functional MR imaging activation, in one subject after visual stimulation (checkerboard, 8 Hz), superimposed on an anatomic image (MP-RAGE). Pixels with a z-score above 4.0 are coded as to the bar on the right. Top row shows runs 1 and 2 from session 1 with system B, middle row shows both runs from session 2 with system A, and bottom row shows the runs from session 3 with system B.

Results

For each of the 12 subjects, six activation maps were created. An example for one subject is shown in Fig 1.

Properties of the Statistical Maps

Motion (as reported by the motion correction algorithm from BrainVoyager) was below 0.2-mm translation and 0.8° rotation for all 72 ETCs (12 subjects, six ETCs per person). The average activated area (with a z-score threshold of 4 and a cluster-size threshold of 250 mm^3) was 38 ± 25 cm^3 (n = 34) for system A and 34 ± 16 cm^3 (n = 38) for system B ($P > .05$). The average signal intensity change was 3.28 ± 0.74% for system A and 2.19 ± 0.44% for system B; this difference was significant ($P < 10^{-3}$).

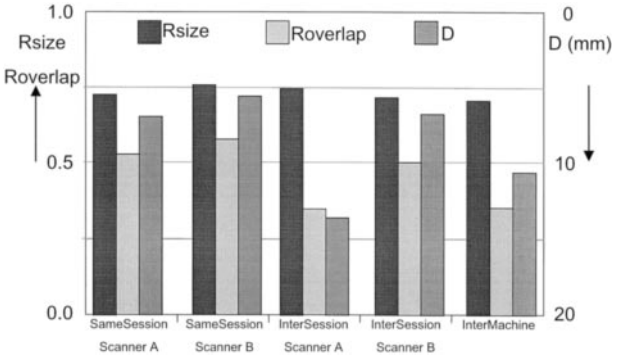

FIG 2. Averages for the reproducibility measures D, R_{size}, and $R_{overlap}$ for same-session and intersession reproducibility for system A and system B, and intermachine reproducibility. D is the distance between the centers of the activated areas, R_{size} is the ratio of the volumes of activated areas, and $R_{overlap}$ is the ratio of the common activated areas. R_{size} and $R_{overlap}$ must be read from the left y axis, and D must be read from the right y axis where the order is inverted.

Reproducibility Values

The averages of D, R_{size}, and $R_{overlap}$ are plotted in Fig 2. R_{size} did not differ significantly between the systems or between the compare conditions (same session, intersession, and intermachine) and was 0.72 ± 0.23 on average.

Results for $R_{overlap}$ and D were as follows: the highest reproducibility values were found in same-session comparisons with $R_{overlap}$ = 0.54 ± 0.20 (mean ± SD) and D = 6 .7 ± 6.6 mm for system A (n = 17); for system B they were $R_{overlap}$ = 0.58 ± 0.17 and D = 5.5 ± 3.6 mm (n = 19). These values did not differ significantly between the systems.

Intersession reproducibility values of system A were $R_{overlap}$ = 0.36 ± 0.19 and D = 13.6 ± 8.5 mm (n = 20), and for system B they were $R_{overlap}$ = 0.51 ± 0.20 and D = 6 .7 ± 5.8 mm (n = 28). Both of these differences were significant, with $P = .008$ and $P < .001$, respectively.

For system A, intersession reproducibility values were significantly lower than same-session reproducibility values ($P = .005$ and $P < .001$, respectively), but for system B, same-session reproducibility values did not differ significantly from intersession reproducibility values.

Intermachine reproducibility values were $R_{overlap}$ = 0.36 ± 0.18 and D = 10.5 ± 6.9 mm (n = 96). These did not differ significantly from the intersession results for system A, but were significantly larger than the intersession results for system B ($P < .001$ and $P < .001$, respectively).

For the above-mentioned results, the z-score threshold was 4.0 and the cluster-size threshold was 250 mm^3. The reproducibility values were also calculated with z-scores ranging from 3 to 6 and cluster-size thresholds from 100 to 1000 mm^3, but whether or not statistical significance was obtained in the above-mentioned comparisons was not influenced by changing these thresholds.

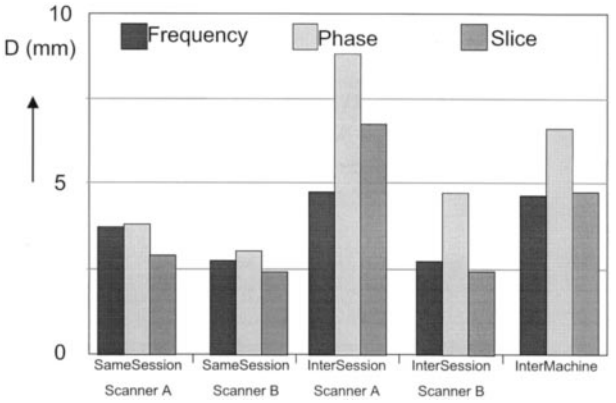

FIG 3. Averages for the reproducibility measure D, separated for the frequency- and phase-encoding directions and the section direction, for same-session and intersession reproducibility for system A and system B, and intermachine reproducibility. D is the distance between the centers of the activated areas.

Influence of Section Positioning and Frequency- and Phase-Encoding Directions

The influence of section positioning and of the frequency- and phase-encoding directions was determined by calculating the (absolute) distance D for the frequency (D_f), phase (D_{ph}), and section (D_s) directions separately. The averages are plotted in Fig 3.

For same-session reproducibility, D_f, D_{ph}, and D_s did not differ significantly, nor did they differ significantly between the two systems, and the average value was 3.1 ± 3.4 mm.

For system A, the intersession D_f was 4.7 ± 6.2 mm, which was significantly smaller than D_{ph} and D_s, which were 8.8 ± 7.2 mm and 6.7 ± 5.0 mm, with the *P* values being .015 and .033, respectively. For system B, the intersession D_{ph}, D_s, and D_f did not differ significantly, the average was 3.3 ± 3.9 mm. The intersession D_f did not differ significantly between the two systems, but D_{ph} and D_s did, the *P* values being .015 and .001, respectively.

For intermachine reproducibility, the values were D_f = 4 .6 ± 4.0 mm, D_{ph} = 6 .6 ± 6.3 mm, and D_s = 4 .7 ± 4.2 mm. None of these differed significantly from the intersession values of system A. Compared with the intersession results of system B, the intermachine D_s was significantly larger (P = .05).

Discussion

If functional MR imaging is to be used for presurgical mapping, it is essential that the results be reproducible. To see whether the reproducibility of functional MR imaging might be MR-system dependent, we performed reproducibility studies on two different MR systems from different manufacturers. This could also be important when functional MR imaging results are exchanged between different centers.

One limitation of this study might be that the two systems produced significantly different signal intensity changes between the rest state and the activated state. However, we found that this did not induce significant differences in activated volume, and therefore any effect on the reproducibility measures presented herein must have been negligible.

Another limitation could be that a small partial-volume effect might have occurred in intersession and intermachine reproducibility; it was only apparent for system A (D_s was significantly larger then D_f). A more subtle section-positioning protocol, such as that proposed by Noll and Eddy (9, 19), might have reduced the partial-volume effect, although we believe it is impossible to avoid this effect completely. The more so because, in the field of surgical mapping, partial-volume effects must be taken into account when determining margins.

A third limitation could be that the method used here to produce statistical maps (Student's *t* test) has been shown to suffer from an inherent lower reproducibility than other methods, for example the so-called slope calculation (13). However, even taking into account the noise in the method used here, we found significant differences between the two systems. Therefore, we expect that the differences between the two systems will be even more significant if more advanced methods for the creation of statistical maps are used.

In the literature, two groups of authors describe visual functional MR imaging reproducibility experiments with the reproducibility measures used here. Rombouts et al (7, 12) performed two reproducibility studies. Their first study measured intersession reproducibility by using only two sections and goggles for the visual stimuli; their second study measured same-session and intersession reproducibility with full-brain coverage. Miki et al (16–18) performed one intersession study by using goggles, and two same-session studies, one with goggles and one with checkerboard stimulation. The results of these five articles together with our results are shown in the Table.

The same-session results from Rombouts et al and

Functional MR Imaging Reproducibility Measures from the Current Study and the Literature

Study	Stimulus	Same Session		Intersession	
		R_{size}	$R_{overlap}$	R_{size}	$R_{overlap}$
Current: system A	Checkerboard	0.73	0.53	0.75	0.38
Current: system B	Checkerboard	0.76	0.58	0.72	0.50
Rombouts et al (7)	Goggles, 2 sections			0.83	0.31
Rombouts et al (12)	Goggles	0.90	0.74	0.88	0.64
Miki et al (16)	Goggles	0.93	0.81		
Miki et al (17)	Checkerboard	0.75	0.61		
Miki et al (18)	Goggles			0.60	0.47

from Miki et al seem to be slightly better than our same-session results. This might be caused by their use of goggles and our use of checkerboard stimuli; in Miki et al (18) it is reported that goggles produced better reproducibility results than those of the checkerboard stimuli. The intersession results from the second study from Rombouts et al (12) are similar to our intersession results with system B. The intersession results from Miki et al are not unlike our intersession results with system A.

The most important findings of the present experiments are twofold. First, same-session reproducibility values were independent of the system used. Second, in intersession reproducibility, the amount of activated voxels (R_{size}) did not depend on the system, but the positions of these voxels ($R_{overlap}$ and D) did: System B reproduced these positions better than did system A. As was to be expected, intermachine reproducibility was roughly the same as the lowest intersession reproducibility, in our case that of system A.

One of the reasons that system A produced lower reproducibility results is that it suffers more from geometric distortions, as can be noted from the much larger D_{ph}. Distortions may have been induced by a less homogeneous magnetic field, perhaps induced by poorer shimming. Another reason for differing results could be that build-in corrections for Nyquist ghost artifacts may differ in effectiveness between the two systems. A robust method to correct both for Nyquist and geometric distortion artifacts is given by Schmithorst et al (22), and a method that reduces geometric distortion artifacts is given by Jenkinson (23). Both methods require off-line reconstruction.

The consequences of our study for multicenter functional MR imaging investigations are limited if the focus of the investigation is on volume effects. If, however, the location is important, intermachine reproducibility might be low and hence could decrease statistical power significantly.

The accuracy of functional MR imaging was previously shown (3, 24, 25) to be usually within 1 cm for more invasive methods such as intraoperative mapping. A consequence of our findings for presurgical planning is that the surgical safety margin should be increased by the distance D reported by intersession reproducibility. In our case, this means that the margin to be added is 13.6 mm for system A and 6.7 mm for system B.

For sites performing functional MR imaging as a tool for presurgical planning, we suggest that intersession reproducibility be determined and that the distance D be added to the surgical safety margin. It is most likely that this distance is smaller when either of the abovementioned methods for correcting Nyquist and/or geometric distortion artifacts is applied.

Other functional MR imaging paradigms than visual ones, including language and motor paradigms, may show other reproducibility data. Intersession reproducibility should therefore be determined for every applied paradigm individually, to obtain a paradigm-dependent surgical safety margin.

Conclusion

Reproducibility of visual functional MR imaging depends on the system used to acquire the functional MR results. The implications for presurgical mapping are that reproducibility measures have to be established per machine. These measures can then be used to determine accurate surgical safety margins.

References

1. Lee CC, Ward HA, Sharbrough FW, et al. **Assessment of functional MR imaging in neurosurgical planning**. *AJNR Am J Neuroradiol* 1999;20:1511–1519
2. Ruge MI, Victor J, Hosain S, et al. **Concordance between functional magnetic resonance imaging and intraoperative language mapping**. *Stereotact Funct Neurosurg* 1999;72:95–102
3. Beisteiner R, Lanzenberger R, Novak K, et al. **Improvement of presurgical patient evaluation by generation of functional magnetic resonance risk maps.** *Neurosci Lett* 2000;290:13–16
4. Hirsch J, Ruge MI, Kim KH, et al. **An integrated functional magnetic resonance imaging procedure for preoperative mapping of cortical areas associated with tactile, motor, language, and visual functions.** *Neurosurgery* 2000;47:711–21, discussion 721–2.
5. Yetkin FZ, McAulie 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
6. 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
7. Rombouts SA, Barkhof F, Hoogenraad FG, 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
8. Genovese CR, Noll DC, Eddy WF. **Estimating test-retest reliability in functional MR imaging. I: statistical methodology.** *Magn Reson Med* 1997;38:497–507
9. Noll DC, Genovese CR, Nystrom LE, et al. **Estimating test-retest reliability in functional MR imaging. II: application to motor and cognitive activation studies.** *Magn Reson Med* 1997;38:508–517
10. Le TH, Hu X. **Methods for assessing accuracy and reliability in functional MRI.** *NMR Biomed* 1997;10:160–164
11. Baumgartner R, Scarth G, Teichtmeister C, Somorjai R, Moser E. **Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. I: reproducibility.** *J Magn Reson Imaging* 1997;7:1094–1101
12. Rombouts SA, Barkhof F, Hoogenraad FG, Sprenger M, Scheltens P. **Within-subject reproducibility of visual activation patterns with functional magnetic resonance imaging using multislice echo planar imaging**. *Magn Reson Imaging* 1998;16:105–113
13. Cohen MS, DuBois RM. **Stability, repeatability, and the expression of signal magnitude in functional magnetic resonance imaging.** *J Magn Reson Imaging* 1999;10:33–40
14. Machielsen WC, Rombouts SA, Barkhof F, Scheltens P, Witter MP. **FMRI of visual encoding: reproducibility of activation.** *Hum Brain Mapp* 2000;9:156–164
15. McGonigle DJ, Howseman AM, Athwal BS, Friston KJ, Frackowiak RS, Holmes AP. **Variability in fMRI: an examination of intersession differences**. *Neuroimage* 2000;11:708–734
16. Miki A, Raz J, van Erp TG, Liu CS, Haselgrove JC, Liu GT. **Reproducibility of visual activation in functional MR imaging and effects of postprocessing**. *AJNR Am J Neuroradiol* 2000;21:910–915
17. Miki A, Raz J, Englander SA, et al. **Reproducibility of visual activation in functional magnetic resonance imaging at very high field strength (4 tesla).** *Jpn J Ophthalmol* 2001;45:1–4
18. Miki A, Liu GT, Englander SA, et al. **Reproducibility of visual activation during checkerboard stimulation in functional magnetic resonance imaging at 4 tesla**. *Jpn J Ophthalmol* 2001;45:151–155
19. Eddy WF, Fitzgerald M, Noll DC. **Improved image registration by using Fourier interpolation.** *Magn Reson Med* 1996;36:923–931
20. Goebel R, Khorram-Sefat D, Muckli L, Hacker H, Singer W. **The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery**. *Eur J Neurosci* 1998;10:1563–1573

21. Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC. **Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold.** *Magn Reson Med* 1995;33:636–647
22. Schmithorst VJ, Dardzinski BJ, Holland SK. **Simultaneous correction of ghost and geometric distortion artifacts in EPI using a multiecho reference scan.** *IEEE Trans Med Imaging* 2001;20:535–539
23. Jenkinson M. **Improved unwarping of EPI volumes using regularised B0 maps**. *Human Brain Mapping* Brighton UK 2001
24. Krings T, Schreckenberger M, Rohde V, et al. **Metabolic and electrophysiological validation of functional MRI**. *J Neurol Neurosurg Psychiatry* 2001;71:762–771
25. Yetkin FZ, Mueller WM, Morris GL, et al. **Functional MR activation correlated with intraoperative cortical mapping.** *AJNR Am J Neuroradiol* 1997;18:1311–1315