NeuroImage **14,** 709–722 (2001)
doi:10.1006/nimg.2001.0869, available online at http://www.idealibrary.com on IDEAL®

Motion Correction Algorithms May Create Spurious Brain Activations in the Absence of Subject Motion

L. Freire*,† and J.-F. Mangin*

**Service Hospitalier Frédéric Joliot, CEA, 91401 Orsay, France; and †Instituto de Medicina Nuclear, FML, 1649-028 Lisbon, Portugal*

Received November 4, 2000

This paper describes several experiments that prove that standard motion correction methods may induce spurious activations in some motion-free fMRI studies. This artifact stems from the fact that activated areas behave like biasing outliers for the difference of square-based measures usually driving such registration methods. This effect is demonstrated first using a motion-free simulated time series including artificial activation-like signal changes. Several additional simulations explore the influence of activation amplitude and extent. The effect is finally highlighted on an actual time series obtained from a 3-T magnet. All the experiments are performed using four different realignment methods, which allows us to show that the problem may be overcome by methods based on a robust similarity measure like mutual information.

Key Words: **fMRI; motion correction; artifact; spurious activation; robust registration.**

INTRODUCTION

Realignment of functional magnetic resonance imaging (fMRI) time series is today considered a required preprocessing step before analysis of functional activation studies. Indeed, when the subject movement is correlated with the task, the changes in signal intensity which arise from head motion can be confused with signal changes due to brain activity (Hajnal *et al.*, 1994). Nevertheless, standard realignment procedures are often not sufficient to correct for all signal changes due to motion. For instance, a nonideal interpolation scheme used to resample realigned images leads to motion-correlated residual intensity errors (Grootoonk *et al.*, 2000). Other motion-correlated residuals may stem from the "spin history effect," which occurs when the spin excitation schedule is changed by the subject motion (Friston *et al.*, 1996; Robson *et al.*, 1997). Finally, other motion-related artifacts can confound fMRI time series, such as intrascan motion and the interaction between motion and susceptibility artifacts (Birn *et al.*, 1997; Wu *et al.*, 1997).

It has been reported that a number of residual motion-related artifacts after realignment are reduced by covarying out signal correlated with functions of the motion estimates (Friston *et al.*, 1996; Grootoonk *et al.*, 2000). It has to be noted, however, that when motion estimates are highly correlated with the task, this regression-based approach is bound to erase some actual activations. While this cost may appear as the price to pay in order to obtain a good protection against false positives, using this approach raises the issue of the motion estimate reliability. Indeed, if ever signal changes induced by the cognitive task slightly bias motion estimates in a systematic task-correlated way, the price of this correction may be very high. Without the correction, however, realignment from task-correlated motion estimates could induce spurious activations. Hence task-correlated motion estimates would be the worst artifact that can be imagined for a realignment method. In this paper, several simulations and some experiments with real data show that this artifact may occur with realignment methods that do not take into account potential outlier voxels related to functional activations when defining the similarity measure which is optimized to assess registration parameters.

A number of papers reporting brain mapping results obtained from fMRI experiments consider the realignment stage of their processing methodology reliable simply because it has been done using one of the standard packages (SPM, Friston *et al.*, 1996; AIR, Woods *et al.*, 1992, 1998). Recurrent difficulties observed in our institution relative to realignment by SPM of time series acquired with our 3-T magnet, however, have led our neuroscientists to follow a rather surprising strategy: compute motion estimates but do not resample time series if motion is small relative to voxel size. This strategy stems from a number of past observations in which predictable activation patterns were obtained without realignment, whereas resampling the time series led to the usual task-correlated motion artifacts (spurious activations along brain edges). When the un-

1053-8119/01 $35.00

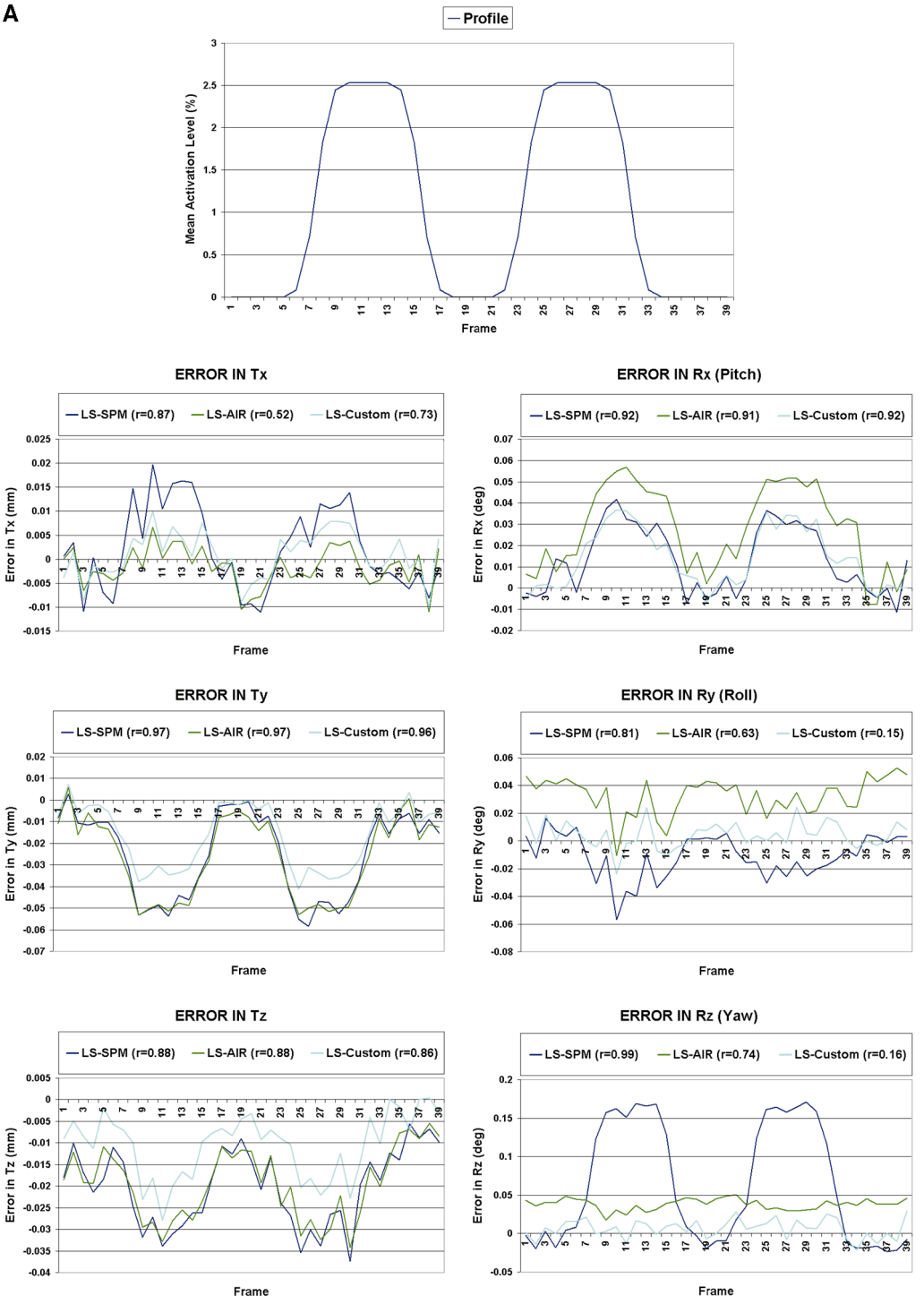


FIG. 1. (A) Artificial activations are added to a motionless constant time series according to the time course given on top. The six realignment parameters are displayed for the LS-based packages. Each parameter time course is cross-correlated with the activation profile. (B) The six realignment parameters displayed for RIU-AIR, MI, and RIU-Custom. RIU-AIR presents no correlation with the time course displayed on the top of A. Small but significant correlations may be observed for RIU-Custom and MI (t_z and *pitch*).

derlying experiments had been designed according to the standard block alternation, some of the estimated motion parameters did actually include an almost perfect periodic trend correlated with the task. The frequency of this problem and the fact that this estimated periodic motion was "too perfect to be honest" led us to suspect a systematic bias of the motion correction algorithm.

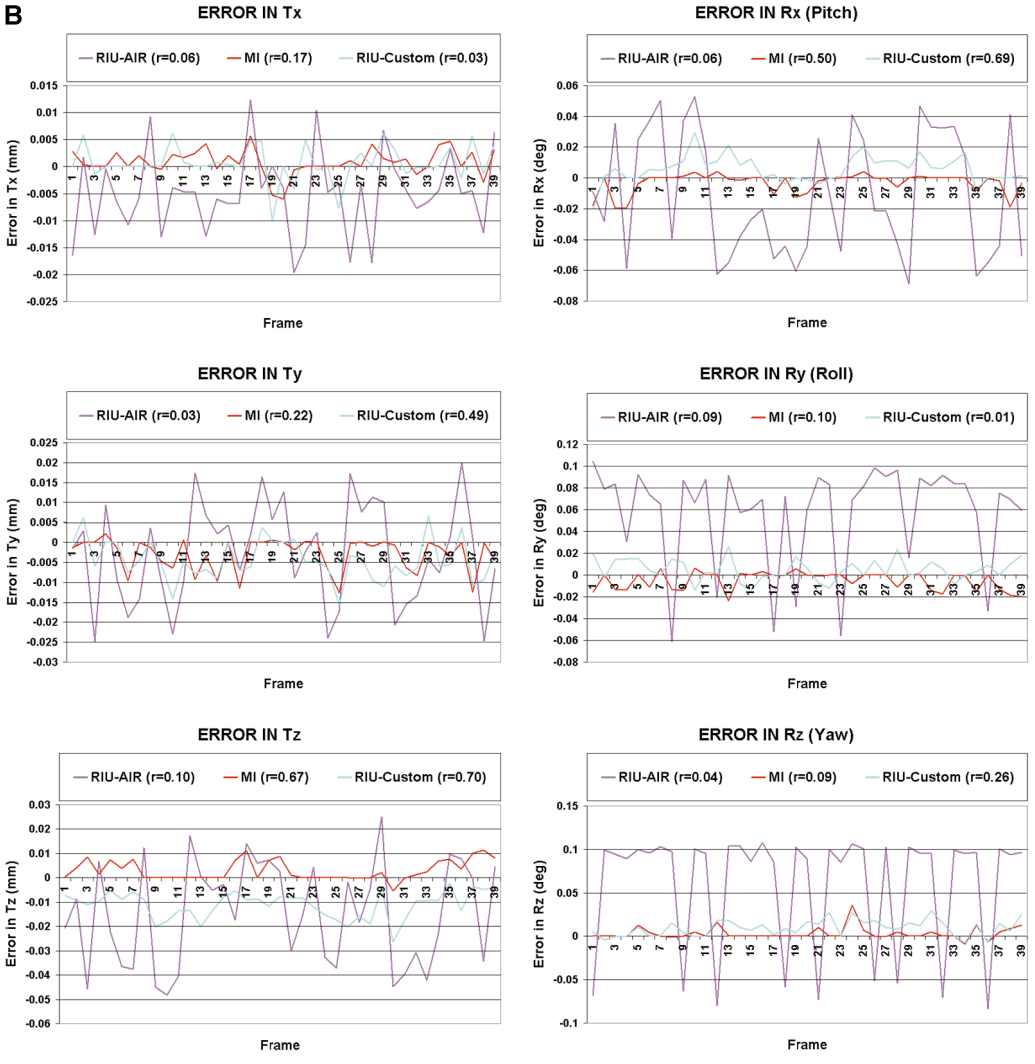
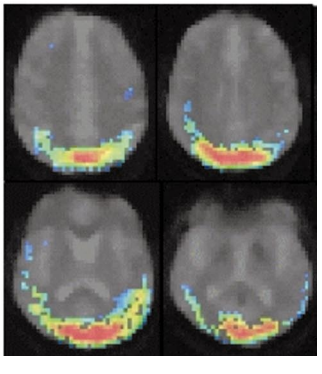


FIG. 1—*Continued*

This paper describes several experiments which demonstrate that this bias is induced by activated areas which behave like outliers for the registration method. The fact that the bias magnitude is highly related to the signal change amplitude may explain why our 3-T magnet led to more difficulties than more usual 1.5-T scanners. Our prediction for the near future, however, is an increasing number of unsatisfied

Pattern	A1	A2	A3
Size (%)	12.4	6.2	3.2
Mean (%)	1.26	1.19	1.18
Max (%)	2.04	2.03	2.03



SCHEME 1. Summary of activation pattern features. Size is given as a percentage of the total number of brain voxels. Mean and Max denote mean and maximum signal increase for the activated voxels.

users of standard motion correction procedures simply related to the widespread used of high-field magnets. Fortunately, our experiments show that more robust similarity measures like "mutual information" could overcome the problem (Wells *et al.*, 1996; Maes *et al.*, 1997; Viola *et al.*, 1997; Studholme *et al.*, 1997; Meyer *et al.*, 1997). It has to be understood that the difficulties with usual measures like the least-squares used in SPM and AIR are not really related to accuracy of the motion estimates. Indeed, reaching a high subvoxel accuracy for motion estimates in actual time series may require better models of the motion-induced signal changes. For instance, spatial distortions related to echo-planar imaging (EPI) depend on the subject position in the scanner, which may confound motion estimation (Jezzard *et al.*, 1999). Therefore, our experiments mainly focus on the potential task-correlated bias observed in motion estimates whatever the estimate actual accuracy.

MATERIAL AND METHODS

fMRI Acquisitions

All fMRI studies were performed on a Brucker scanner operating at 3-T using a 30-contiguous-slice 2D EPI sequence (slice array of 64 $\times$ 64 voxels). This sequence had in-plane resolution of 3.75 mm and slice thickness of 4 mm. The potential bias induced by activations in realignment algorithms was evaluated in a human study using a design of two alternating visual stimuli. The subject's head was cushioned inside the Brucker proprietary head *rf* coil assembly, and two adjustable pads exerted light pressure to either side of the head.

Similarity Measures

To bring two images into spatial alignment, a geometrical transformation is applied to one of the images. The type of the transformation can range from a simple rigid body transformation to a fully elastic transformation. All the experiments described in this paper have been performed using a rigid body transformation, which is described by three translation and three rotation parameters. The purpose of a similarity measure is to return a value indicating how well two images match given a certain transformation. Ideally, by maximizing the similarity measure one should find the transformation that registers the images. The optimal rigid body transformation, however, usually depends on the chosen similarity measure and on the implementation of its optimization. The goal of this paper is to assess at which extent the optimal transformations given by standard realignment methods are biased by activations.

Four different realignment methods are used in all our experiments. Each underlying implementation depends on a few parameters, which may slightly modify the realignment results. A number of works have been dedicated to evaluation of registration methods accuracy (Jiang *et al.*, 1995; Frouin *et al.*, 1997; West *et al.*, 1997; Woods *et al.*, 1998; Holden *et al.*, 2000). While this is clearly a key point to compare similarity measures, such work requires the study of each parameter influence, which is far beyond the scope of this paper. Since our main goal is to highlight the potential bias induced by activations, we have chosen to set each parameter either to the best choice leading to acceptable computation time or to the value commonly used by standard users:

(1) *LS-SPM:* the standard realignment algorithm in SPM96 (http://www.fil.ion.ucl.ac.uk/spm; Friston *et al.*, 1995). The underlying similarity measure is simply a least-square, namely the sum of the squared discrepancies between both images. One specificity of SPM implementation is the use of a first-order Taylor series approximation of the rigid body transformation effects. While this choice allows rapid minimization of the measure iteratively using a singular value decomposition (SVD), it may explain some differences with other implementations of the same similarity measure (Hill *et al.*, 2001). A 3D smoothing is applied before realignment to ensure a good behavior of the Taylor expansion. We set the Gaussian kernel full width at half-maximum (FWHM) to 8 mm. The number of iterations was set to 16. The whole set of slices was included in the SVD. We have checked that the realignment algorithm in SPM99, while slightly different (points outside the head are removed from the SVD), presents the same qualitative behavior relative to the bias induced by activations.

(2) *LS-AIR:* a second implementation of the least-square approach available in AIR 2.0 (Woods *et al.*, 1992). The minimization is done according to a Powell-like unidimensional algorithm. A 3D Gaussian smoothing is applied to the data to get more robust minimization (FWHM = 4 mm). A threshold was used to discard voxels with low signal (mainly located outside the head).

(3) *RIU-AIR:* the ratio image uniformity similarity function of AIR 2.0 (Woods *et al.*, 1992). This function is simply the standard deviation of a ratio image computed on a voxel-by-voxel basis. Minimization of this cost function increases the uniformity of the ratio image, which is independent of global scaling of the original images, and improves registration. Preprocessing and minimization are performed as for the previous LS-AIR measure. The measures used in AIR 3.0 are the same but the minimization implementation has been refined.

(4) *MI:* mutual information (Wells *et al.*, 1996; Maes *et al.*, 1997; Viola *et al.*, 1997; Studholme *et al.*, 1997; Meyer *et al.*, 1997). Mutual information is a measure originating from information theory, which is today

accepted by many as one of the most accurate and robust registration measures. The underlying concept is entropy. The entropy of an image can be thought of as a measure of dispersion in the distribution of the image gray values. Given two images A and B, the definition of the mutual information MI(A,B) of these images is MI(A,B) = E(A) + E(B) − E(A,B), with E(A) and E(B) the entropies of the images A and B, respectively, and E(A,B) their joint entropy. The joint entropy E(A,B) measures the dispersion of the joint probability distribution $p(a,b)$: the probability of the occurrence of gray value a in image A and gray value b in image B (at the same position), for all a and b in the overlapping part of A and B. The joint probability distribution should have fewer and sharper peaks when the images are matched than for any case of misalignment. Therefore maximization of mutual information should correspond to the best registration. An implementation of MI can be found in SPM99. For the experiments described in this paper, however, our own implementation was used. Minimization was performed according to a Powell-like unidimensional scheme. Image resampling was achieved using cubic spline interpolation. The joint histogram computation includes a rebinning of each image gray level set to 64 values, of which each voxel gray level contributes to the two closest values proportionally to the two underlying intervals.

Two additional methods are used in our first experiment:

(5) *LS-CUSTOM:* a third implementation of the difference of square measure, which optimization is performed according to the same scheme as for our MI-based method (cubic spline interpolation, Powell algorithm) after an 8-mm Gaussian smoothing equivalent to the SPM96 one.

(6) *RIU-CUSTOM:* a second implementation of the ratio image uniformity measure, which optimization is performed according to the same scheme as for our MI-based method after a 4-mm Gaussian smoothing equivalent to the AIR 2.0 one.

These custom implementations of SPM and AIR similarity measures are used to discard some potential confound related to our MI custom implementation. For instance, these implementations allows us to study potential artifacts related to the fact that simulated time series stem from a cubic spline interpolation, while SPM and AIR use other interpolation schemes (truncated *sinc* and linear interpolation), which may result in different behaviors of the three LS-based methods. LS-CUSTOM and RIU-CUSTOM methods will be used only during the first experiment.

Simulations

Evaluation of the putative biasing effect due to activations was first achieved using artificial time series. Each volume in the time series was created by applying an artificial rigid-body motion T_{sim} to a reference image using a cubic spline-based interpolation method available on the World Wide Web: http://bigwww.epfl.ch/algorithms.html (Unser *et al.*, 1993a,b). This method embeds the volume in a surrounding space filled with null value. The reference image (64 × 64 × 30, 3.75 × 3.75 × 4 mm) was one of the EPI BOLD images of the study mentioned above denoised with a standard 3 × 3 × 3 median filter. Gaussian noise was added to the reference image and to all frames of the time series in order to simulate the effects of thermal noise in fMRI scans (standard deviation 2.5% of mean cerebral voxel value). Various artificial activations were then added either to the reference image or to the rest of the time series according to the simulation requirement. Three different activation patterns were manually drawn in the occipital lobe in order to mimic some visual activations observed during the underlying neuroscience study. These patterns were first filled with a random noise and then spatially filtered with a Gaussian (standard deviation 2 mm). The resulting image was then masked according to the initial pattern. Some features of the final patterns are summarized in Scheme 1. A few slices presented in Scheme 1 give an idea on the A1 activation spatial profile. A range of activation amplitudes was studied using multiplicative factors.

Each frame of the artificial time series is aligned to the reference image using one of the registration methods, which yields an estimated rigid-body transformation T_{est}. Hence the alignment error is given by the residual rigid-body transformation $T_{res} = T_{sim} \times {T_{est}}^{-1}$, where each transformation is represented by a standard homogeneous matrix. The origin for rotations is located in the center of the volume. The translation (E_t) and rotation (E_r) alignment errors are given by $E_t = sqrt(T(1,4)^2 + T(2,4)^2 + T(3,4)^2)$ (in mm) and $E_r = \cos^{-1}[(T(1,1) + T(2,2) + T(3,3) - 1)/2]$ (in degrees). When required, the six motion parameters of a transformation T are given by: $t_x = T(1,4)$, $t_y = T(2,4)$, $t_z = T(3,4)$, $r_y = \sin^{-1}(T(1,3))$, $r_x = \sin^{-1}(T(2,3)/\cos(r_y))$, $r_z = \sin^{-1}(T(1,2)/\cos(r_y))$.

EXPERIMENTS

Simulated Activations without Motion

The first experiment investigates whether some realignment method may lead to artifactual task-related motion estimates in the absence of any initial misalignment in the time series. A second issue is whether motion estimates biased by actual activations may induce additional spurious activations. The different steps of this experiment can be summarized as follows:

- Generate an artificial time series by duplicating the reference image 40 times.
- Include in each frame the activation pattern A1 (see Scheme 1) multiplied by an intensity which varies

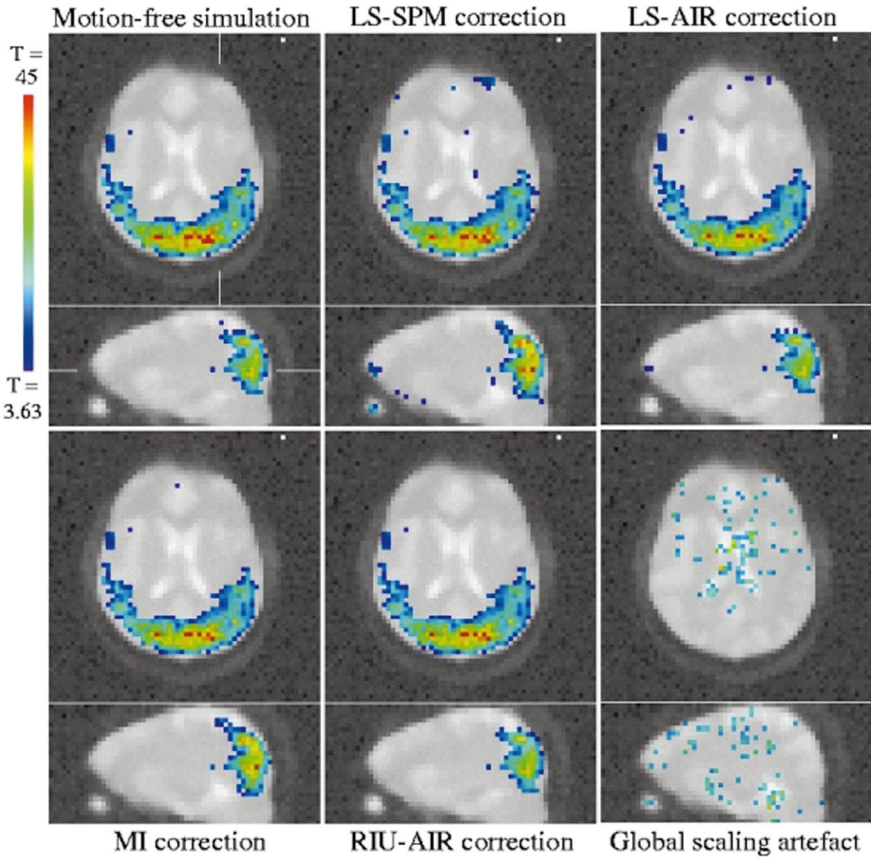


FIG. 2. One axial and one sagittal slice of the activation maps obtained from SPM99 after using the different registration methods. Spurious activated voxels can be observed when using LS-based methods.

throughout the time series according to the time course given in Fig. 1 (two square stimuli convolved with a simple hemodynamic response; the maximal mean activation is 2.52%).

- Run the six registration methods.
- Evaluate the six transformation parameters of T_{est} for each package (see Fig. 1).
- Compute cross-correlation between each parameter and the A1 time course (see Fig. 1).
- Infer activated areas from the realigned time series issued from the four first registration methods using SPM99 (see Fig. 2).

Several realignment parameters related to the least-square-based methods (LS-SPM, LS-AIR, and LS-Custom) demonstrate a high correlation with the time course of the simulated activation (see Fig. 1A): for LS-SPM, the highest correlation is obtained for the *yaw* parameter (0.99); for LS-AIR and LS-Custom the maximum correlations are obtained for the t_y parameter (0.97 and 0.96, respectively). The highest amplitude of the task-related parameter time course is 0.05 mm (t_y) and 0.15° (*yaw*) for LS-SPM, 0.05 mm (t_y) and 0.04° (*pitch*) for LS-AIR, and 0.04 (t_y) and 0.04° (*pitch*) for LS-Custom. Lower but significant correlations are observed for some parameters related to MI and RIU-Custom (0.67 and 0.70, respectively, for t_z), but amplitude of the task-related time course is smaller: 0.01 mm (t_z) and 0.02° (*pitch*) for MI, 0.02 mm (t_z) and 0.03° (*pitch*) for RIU-Custom. Finally, no significant correlation is observed for RIU-AIR (0.10 for the highest one), but the realignment curves include more noise than for the other methods (see Fig. 1B).

The initial time series was realigned from each of the four first motion estimations using a cubic-spline interpolation. The generalized linear model was then used to fit each voxel with the artificial profile of Fig. 1 using SPM99 (http://www.fil.ion.ucl.ac.uk/spm) after the following standard preprocessing: spatial Gaussian smoothing (full-width at half-maximum 5 mm) and low-pass temporal filtering by a Gaussian function with a two-frame width. The voxels were reported as activated if the *P* value exceeded a threshold of 0.001 uncorrected for multiple comparisons. In order to quantify the false positive rate, we have computed the number of activated voxels located outside a dilated version of the activated area inferred without motion correction (3148 activated voxels). The dilation of a size 1 voxel allows us to get rid of a potential confound

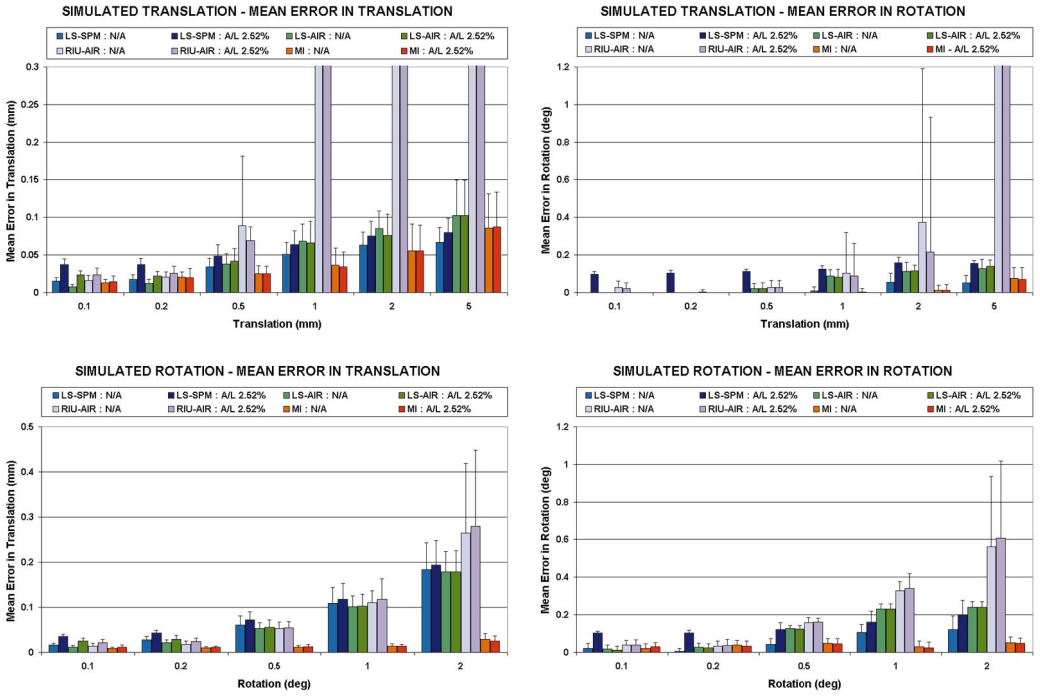


FIG. 3. Influence of activation on registration accuracy for a wide range of simulated misregistrations. For each method, accuracy in the situation of no activation (N/A, lighter color) is compared with accuracy when 12.4% of the brain (occipital area, A1) is activated in the reference volume with mean signal increase of 2.52% (A, darker color). Charts refer to simulated translation (top) and simulated rotation (bottom) experiments and produce means and standard deviations of E_t (left) and E_r (right).

related to threshold effects around the simulated activation. Two hundred twenty-seven spurious voxels have been observed for LS-SPM, 16 spurious voxels for LS-AIR, and no spurious voxels for MI and RIU-AIR.

An illustration of the consequence of the activation-correlated motion estimates is proposed for a slice of the brain in Fig. 2. Spurious activated voxels can be observed along brain edges after LS-SPM and LS-AIR motion corrections. While spurious activated voxels were not individually significant after correction for multiple comparisons, the fact that they were gathered along brain edges led to significant clusters for LS-SPM, for which 6 spurious activated clusters with an extent exceeding a threshold of $P = 0.05$ corrected for multiple comparison were found. The largest significant cluster was made up of 22 voxels. In return no significant cluster was observed for LS-AIR.

In order to illustrate another potential artifact induced by the presence of activations, we performed a second study from the initial time series. Each frame was divided by its mean value before fitting the artificial profile. This approach is sometimes used to discard global scaling effects related to MR acquisitions (Andersson *et al.*, 1997). Here, because of the bias induced on frame mean values by the presence of activations, a lot of voxels turned out to be anticorrelated with the artificial profile ($P < 0.001$ uncorrected) (see Fig. 2—global scaling artifact).

Simulated Activations with Motion

The second experiment investigates influence of activations on registration method accuracy. A method robust to the presence of activations in the time series should keep the same level of accuracy whatever the activation features. The important point here is not the absolute accuracy of the method, which could depend on the tuning of some intrinsic parameters, but the potential accuracy weakening induced by signal change in activated areas relative to the reference image. This experiment relies on a huge number of simulated volumes, which allows us to study the influence of several parameters on a statistical basis. In order to get rid of potential bias related to field-of-view variations after simulated motion, all volumes were stripped from their border voxels before realignment in order to reach a $62 \times 62 \times 28$ geometry (a subvolume was used). To eliminate simulated motion specificities relative to the reference volume axes as a potential confound, the simulated translations were applied systematically in the 20 directions of a regular dodecahedron, and the simulated rotations were applied around the 20 different axes defined by the same dodecahedron. Hence, for a given translation or rotation amplitude, accuracy was assessed from means and standard deviations of translation (E_t) and rotation (E_r) errors relative to 20 different realignments.

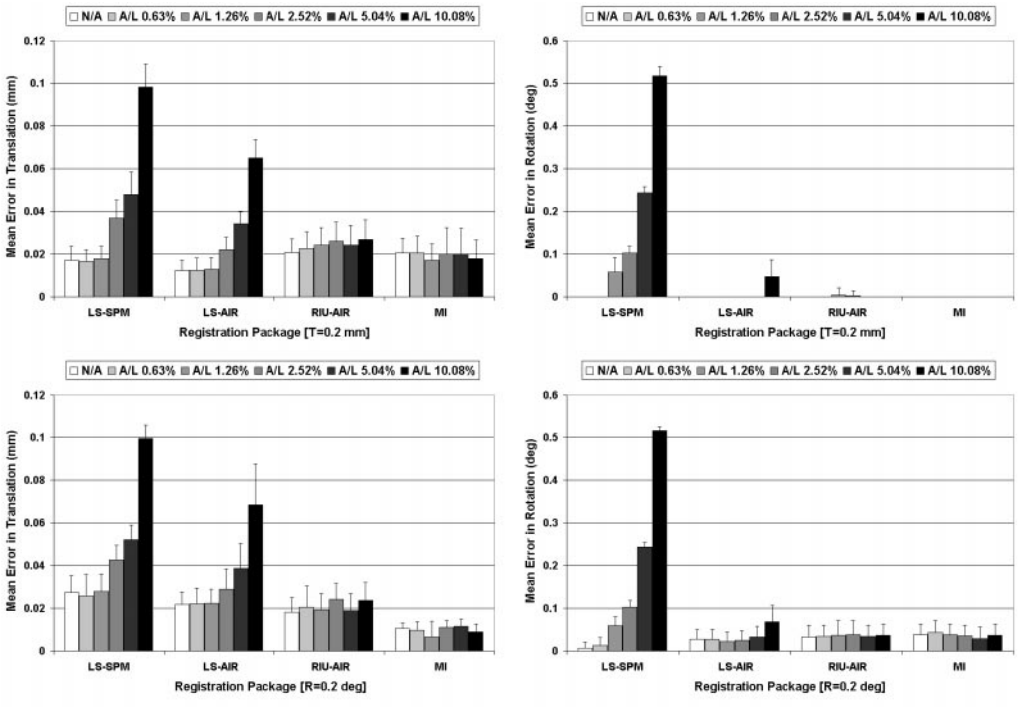


FIG. 4. Influence of activation intensity on registration accuracy. Accuracy of the different registration methods is evaluated by means of mean and standard deviation values of E_t (left) and E_r (right) for increasing activation intensities. 12.4% of the brain (occipital area, A1) is activated in the reference volume with mean signal increase ranging from 0.63 to 10.08% and results are displayed with no-activation situation (N/A) for comparison. Charts refer to simulated misregistrations of 0.2 mm (top) and 0.2° (bottom).

To study influence of motion amplitude, 11 time series of 20 volumes were generated according to the strategy mentioned above for six translation amplitudes (0.1, 0.2, 0.5, 1.0, 2.0, and 5.0 mm) and five rotation amplitudes (0.1, 0.2, 0.5, 1.0, and 2.0°). Each time series was realigned with eight different reference images: the activation-free version used initially to compute the series and seven modified versions including activated areas (A1 with 0.63, 1.26, 2.52, 5.04, and 10.08% mean signal increase and A2 and A3 with 2.52% mean signal increase). Realignment was performed with the four first methods.

Motion Amplitude

The influence of motion amplitude on the putative activation related bias is studied first. For each method and each motion amplitude, accuracy without activation is compared with accuracy when 12.4% of the brain (occipital area, A1) is activated in the reference volume with mean signal increase of 2.52% (see Fig. 3). In all cases, activations produce significant decline of LS-SPM accuracy, whereas this effect is restricted to the translation error (E_t) and the smallest translations for LS-AIR. A less significant but similar effect is observed for RIU-AIR. In return MI accuracy does not depend on activation. The large errors related to RIU-AIR in some situations result from some optimization problems. The values obtained are meaningless and their representation in the corresponding charts would reduce the *y*-axis scale.

Activation Level and Size

The influence of activation level was studied for 0.2 mm translations and 0.2° rotations (Fig. 4). A1 pattern was added to reference image with 0.63, 1.26, 2.52, 5.04, and 10.08% mean signal increase and the results were compared with the situation of no activation. LS-SPM accuracy declines linearly relative to activation mean signal increase whatever the considered error (E_t and E_r). A similar but smaller effect is observed for LS-AIR, especially for E_t. The accuracy of the two other methods does not depend on the signal increase amplitude in the activated area.

The influence of activation size was also studied for translations of 0.2 mm and rotations of 0.2° (Fig. 5). A1, A2, and A3 patterns were added separately to reference image (with 2.52% mean signal increase in activated regions) and results displayed with the nonactivated situation for comparison. LS-SPM accuracy declines significantly when the activated area is enlarged, while this effect is smaller for AIR methods and nonexistent for MI. When an LS-based method is used, activation level has a more dramatic role on the accuracy decline than activation size. When activated region size is doubled, LS-SPM errors in rotation (the most modified ones) are typically increased by a factor

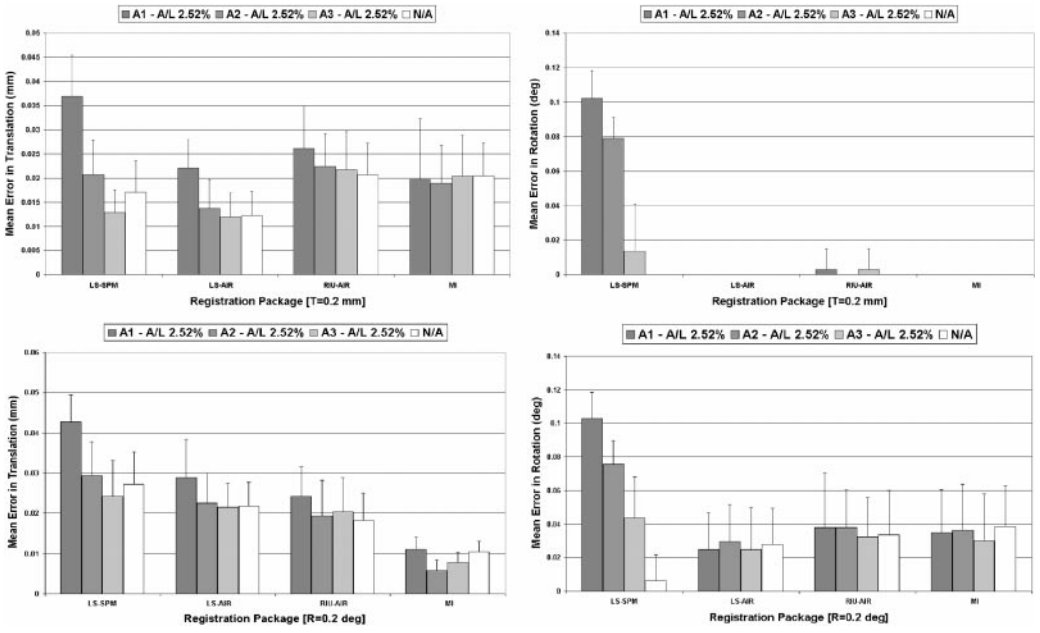


FIG. 5. Influence of activation size on registration accuracy. For each method, accuracy with decreasing activation size is displayed when 12.4, 6.2, and 3.2% of the brain (occipital area, A1, A2, and A3) is activated in the reference volume with mean signal increase of 2.52%. N/A refers to nonactivated situation. Charts produce means and standard deviations of E_t (left) and E_r (right) and refer to simulated misregistrations of 0.2 mm (top) and 0.2° (bottom).

of 25 to 50%. When activation level is doubled, however, LS-SPM errors in rotation may be doubled.

Experiments with Actual Time Series

Finally, the first four registration methods were run on an actual time series made up of 180 frames. The repeated stimulus period corresponds to 18 frames (2 s acquisition per frame). Each period alternates two 9-frame-long presentations of two cognitively different visual stimuli. The six rigid-body registration parameters are displayed in Fig. 6 for the first four registration packages. The general trends of the six parameters' estimations are consistent across methods apart from the *yaw* parameter. It should be noted that according to the estimation results, the actual motion amplitude was rather small (less than 0.15° and 0.15 mm for all frames). Some of the charts clearly display stimulus-correlated periodic variations. The highest correlations with a simple model of the hemodynamic answer in activated area are 0.79 for LS-SPM, 0.64 for LS-AIR, 0.55 for RIU-AIR, and 0.46 for MI. The more impressive periodic effect is observed on the *pitch* chart for LS-SPM and LS-AIR, while this periodic trend is less clear for RIU-AIR and MI (see Fig. 7 for more detailed examination).

Like in the first experiment, the actual time series was realigned from each of the four motion estimations using a cubic spline interpolation. SPM99 was used then to perform detection of activations. The following standard preprocessing was applied: spatial Gaussian smoothing (full-width at half-maximum 5 mm), high-pass temporal filtering (period 120 s), and low-pass temporal filtering by a Gaussian function with a 4-s width. The generalized linear model was used then to fit each voxel with a linear combination of two functions: the first was derived by convolving a standard hemodynamic response function with the periodic stimulus, the second was the time derivative of the first in order to model possible variations in activation onset. The voxels were reported as activated if the *P* value exceeded a threshold of 0.05 corrected for multiple comparisons.

An illustration of the consequences of the stimulus-correlated motion estimates is shown for a few slices of the brain in Fig. 8. Considering the activation map obtained from the raw time series as a reference, a number of additional activated voxels are observed along some high-contrast brain edges after LS-SPM motion correction and to a smaller extent after LS-AIR correction. RIU-AIR and MI corrections have a very small influence on the activation map. The effect related to LS-SPM correction has been observed for numerous cognitive experiments in our institution.

DISCUSSION

All retrospective image registration algorithms rely on a similarity measure, which has to be maximized in order to achieve the result. A huge number of different measures have been proposed in the literature (Maintz

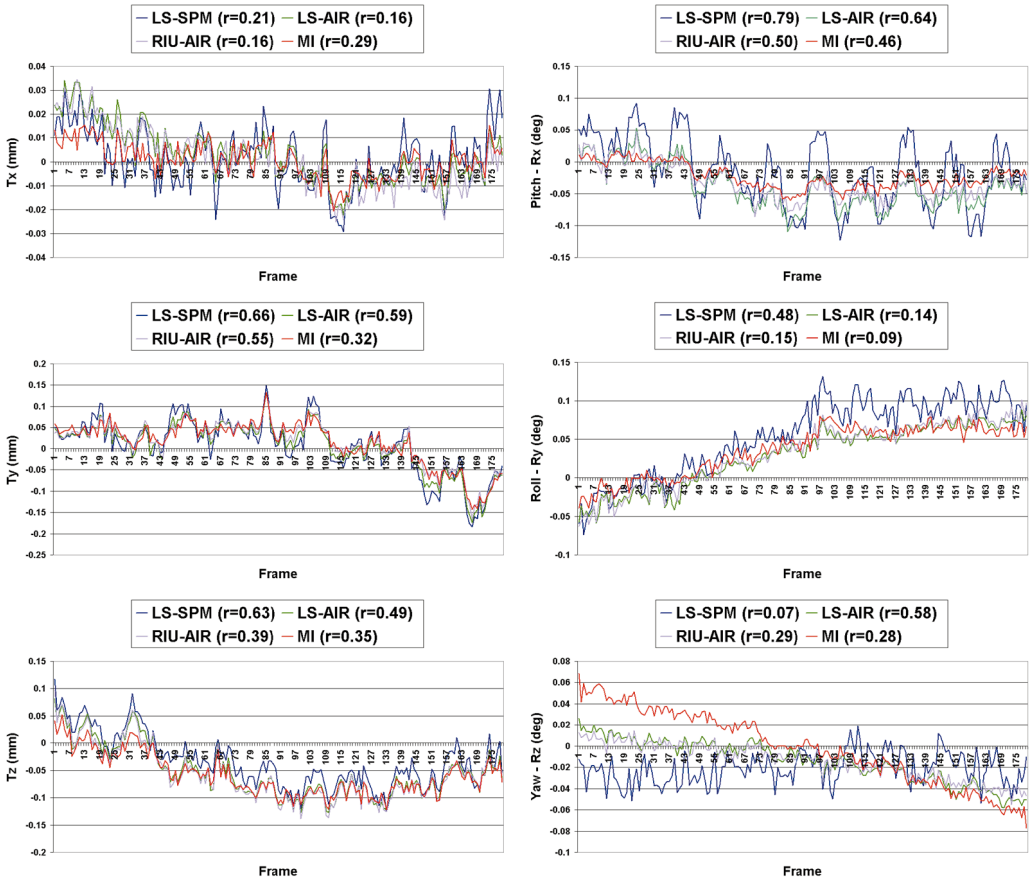


FIG. 6. Motion correction parameters for the four registration methods for an actual time series of 180 frames. The underlying stimulus is made up of 10 periods of 18 frames, each period consisting of two alternating 9-frame-long blocks with a different visual presentation. Stimulus-correlated periodic trends can be observed on some of the charts, especially for least-square-based methods. The correlation with a simple model of the hemodynamic answer to the task is given for each parameter.

et al., 1998; Hill *et al.*, 2001). One important feature helping to distinguish two classes of similarity measures is the robustness to potential outliers, namely voxels that do not verify some of the assumptions underlying the measure design. Robust measures have been classically proposed to register multimodal images, while simpler difference of square-based measures are usually employed for time-series motion correction. The experiments performed in this paper tend to prove that this choice may be questioned because of the presence of activated areas in standard fMRI time series. Indeed, least-square-based approaches are known to be highly sensitive to such outliers (Meer *et al.*, 1991).

The first simulation has shown that LS-SPM, LS-AIR, and LS-Custom motion parameter estimations

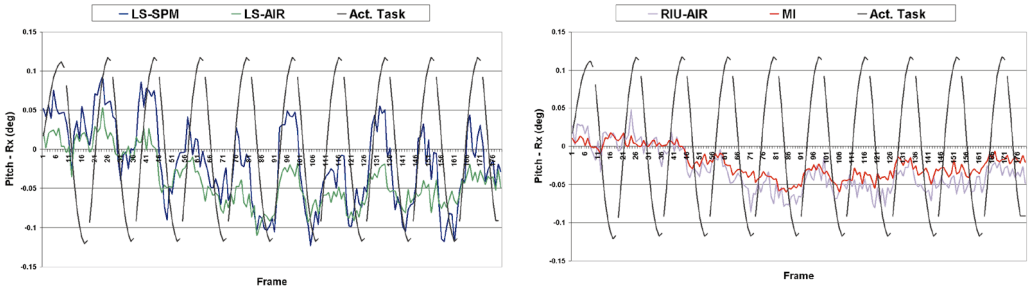


FIG. 7. Isolated charts of *pitch* curves displayed in Fig. 6. Left: LS methods. Right: RIU-AIR and MI. From the LS-SPM curve it is possible to infer the periodicity of the stimulation paradigm, which corresponds to 18 frames/cycle. This gives an idea about the strong correlation between the activation and the registration parameters when an LS-based method is used. MI and RIU-AIR curves present a much less correlated behavior.

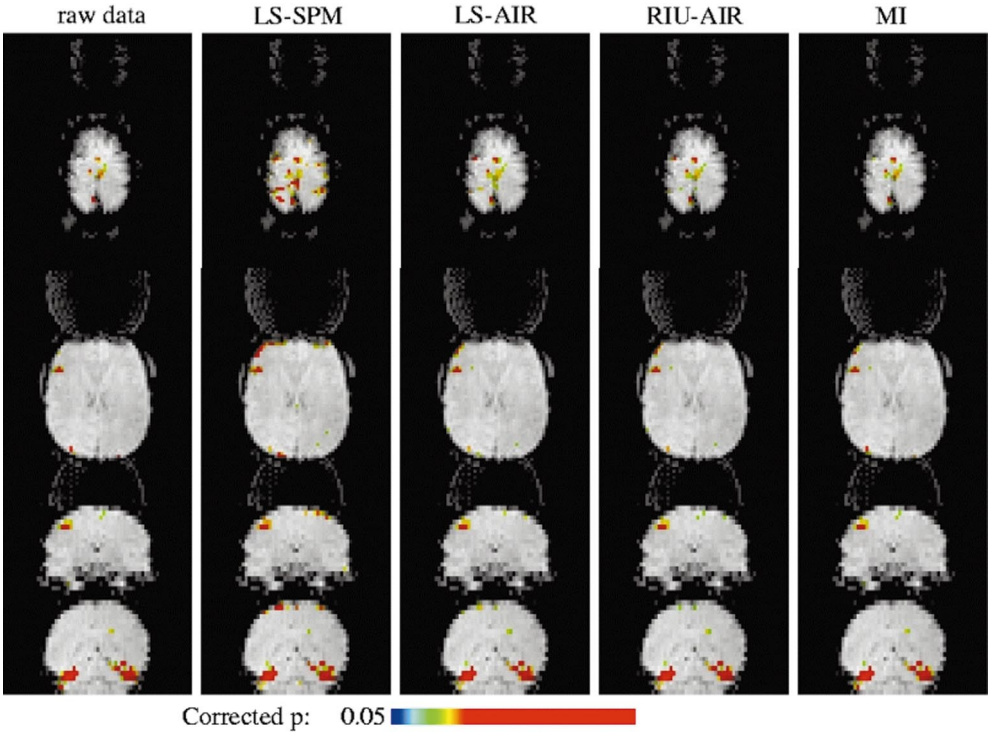


FIG. 8. A few slices of the activation maps obtained from SPM99 after realignment using the different methods.

are biased by signal changes related to activated areas. The fact that the three different implementations of the difference of square measure are biased tends to prove that this bias is related to the nature of this measure. Difference of squares, indeed, is an optimal estimator when residuals are endowed with a Gaussian distribution, which is not verified for the motion correction application because of the presence of activated areas. Furthermore, this first experiment has proved that this bias may induce spurious activations along high-contrast brain edges during the following data analysis. Of course, some of the features of this simulation may be discussed as unrealistic (activation level and size, noise model, no spatial distortions, etc.). For instance, usual activation amplitudes observed in the visual system are in the range tested in this paper, while complex vascular effects can create even larger amplitudes up to 20% (Turner *et al.*, 1993). Our simulation, however, highlights a weakness of the difference of square-based measures that may be overcome by more robust measures. The fact that LS-SPM motion correction led to the apparition of spurious activated clusters with a large extent, indeed, is especially disturbing. While LS-AIR seems less sensitive, this simulation has shown that it is not bias proof.

While almost insensitive to activations in this simulation, the two other methods have presented two qualitatively different behaviors. The RIU-AIR method seems to lead to local maxima difficulties (perhaps related to a bad tuning of the method implementation during our experiment). This results in a low accuracy, which hides any potential activation-related bias. The results obtained in experiment 1 from another implementation of RIU have, indeed, shown that this similarity measure is not bias proof. MI has presented the best behavior with a very small bias amplitude without important influence on the activation detection process. Of course, this simulation does not prove that MI would have a correct behavior in any situation.

The behavior of the RIU-AIR method during the first experiment highlights an important point to be understood. The problem induced by activation-related bias is not related to actual accuracy. For instance, corrupting LS-SPM motion parameter estimations with a reasonable random noise may be sufficient to get rid of spurious activations while preserving actual ones (losing some statistical sensitivity). This observation is illustrated by the results of the second experiment in which the activation influence on registration accuracy is significant only for small motions. Indeed, larger motions lead to a lower registration accuracy, which masks the activation-related bias. This could explain the surprising heuristics of our institution's neuroscientists, which discard realignment only for small amplitude estimated motions.

The second set of simulations shows that RIU-AIR and MI accuracy do not depend on the presence of activations. It should be noted that the very high accuracy of MI may result in part from the implementation of its optimization, which relies on the cubic spline

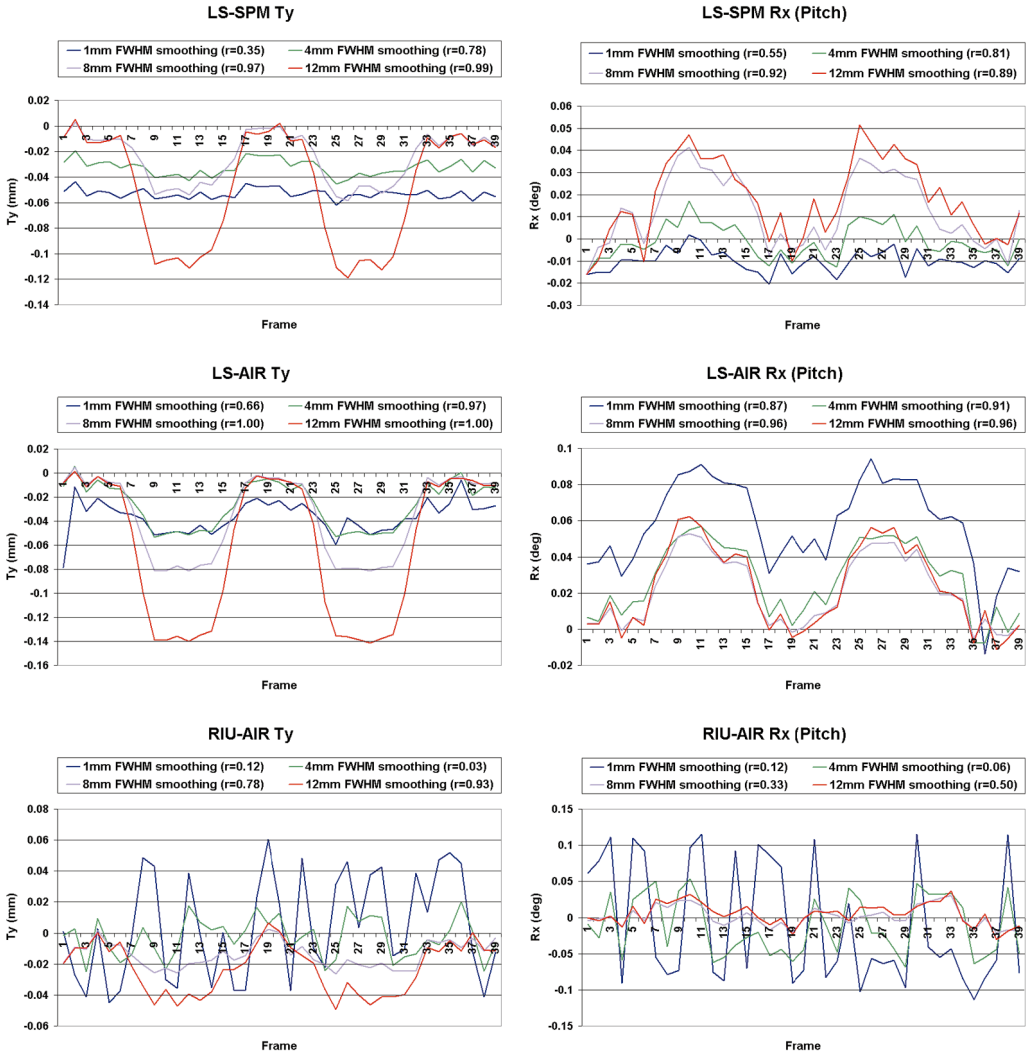


FIG. 9. Influence of spatial smoothing on activation-related bias for SPM96 and AIR 2.0 methods. Top, LS-SPM; middle, LS-AIR; bottom, RIU-AIR.

interpolator used to father the simulated time series. Hence, absolute accuracies should not be compared across methods. The RIU-AIR method, however, seems more prone to problematic optimization, which explains its poor mean accuracy. Surprisingly, this sensitivity to local minima may be a good protection against the bias, because estimated motion parameter variance is greater than any potential underlying bias.

While our point of view during this paper was the study of registration packages used with the usual default parameters, we have finally performed an additional experiment on the influence of the initial spatial smoothing applied by SPM and AIR. This smoothing is supposed to reduce the number of local maxima of the similarity measure and allows SPM to use a reliable Taylor expansion. Our initial guess was a high influence of this smoothing on the bias amplitude, because the smoothing was bound to increase the outlier nature of activated areas. Without any smoothing, indeed, activation amplitude is at the noise level. Hence, we have performed experiment 1 with SPM and AIR methods after different levels of smoothing. Figure 9 shows the resulting charts for a few motion parameters. These results confirm our guess. For instance, RIU-AIR optimization problems are overcome by a large smoothing, but motion estimates turn out to be biased. LS-SPM-related charts show that without smoothing, the bias is very low but motion estimates are less accurate than after a smoothing, which is related to Taylor expansion.

The simulations dedicated to the study of activation size and amplitude influence give a good explanation for the fact that the problem seems more frequent on our 3-T magnet than at other institutions. As it could be predicted for least squares, the bias amplitude is increased according to the signal change amplitude (Meer *et al.*, 1991). Furthermore, the fact that the bias amplitude depends on the activation size explains why

the problem is dependent on the cognitive experiment. One could predict that the bias is also largely dependent on the activation location.

Our experiment with actual time series seems to be consistent with our interpretation of the simulation studies. The arguments that lead us to discard actual task-correlated motion during data acquisition are the following: the periodic motion amplitude estimated by LS-SPM and LS-AIR on the *pitch* chart is different. Moreover, the two other methods do not detect this putative motion. Finally, this periodic motion amplitude is approximately the same for each stimulus period, which would be rather surprising for an actual motion. The fact that all methods do not agree on the estimated *yaw* parameter is of course very difficult to understand. One possible explanation could stem from the fact that the rigid-body transformation is not sufficient to correct for all the consequences of the motion because of distortions. The discord on the periodic motion, however, seems of a different nature and leads to alarming effects on activation maps.

If our interpretation is correct, LS-SPM correction, and to a smaller extent LS-AIR, creates spurious clusters of activated voxels along high-contrast brain edges. In our opinion, the localization of these spurious clusters depends only on the brain edge orientation relative to the actual activation localization. This could mean that spurious activations may appear at the same place across individuals performing in the same cognitive experiment and hence survive to group analysis. While we hope that this alarming prediction is too pessimistic, it calls for trying to minimize the risk.

Our work has shown that more sophisticated similarity measures like MI could clarify the situation thanks to their robustness to outliers. While MI was used for historical reasons during our work, this may not be the best choice for motion correction, first because of computational time considerations, second because recent results have shown that MI is prone to local maxima problems (Roche *et al.*, 1998; Studholme *et al.*, 1999; Pluim *et al.*, 2000a). While RIU-AIR may appear as an alternative at first glance, its nonconvexity problems seem worse than for MI and could be bothersome in case of large motions. Moreover, RIU similarity measure seems potentially biased in some situations according to our experiment with RIU-Custom and the level of spatial smoothing. In fact, the field of robust similarity measure is currently very active and should provide other adequate solutions (Nikou *et al.*, 1998; Roche *et al.*, 1998, 2000; Pluim *et al.*, 2000b; Hill *et al.*, 2001). Further work has to be done, however, to study the behavior of the different measures relative to activations. Activations, indeed, induce only faint signal modifications to be compared, for instance, to pathological or postsurgery modifications. Hence, the choice of the best cost functions requires further experiments with various activation paradigms and different magnets. We hope that our observation will trigger such experiments.

ACKNOWLEDGMENTS

We thank J.-B. Poline, S. Berthoz, P. F. Van de Moortele, S. Dehaene, D. LeBihan, and V. Frouin for the stimulating discussions about motion correction-related problems.

REFERENCES

Andersson, J. L. R. 1997. How to estimate global activity independent of changes in local activity. *NeuroImage* **6:** 237–244.

Birn, R. M., Jesmanowicz, A., Cor, R., and Shaker, R. 1997. Correction of dynamic Bz-field artifacts in EPI. In *Proc. ISMRM 5th Annual Meeting, Vancouver,* p. 1913.

Friston, K. J., Ashburner, J., Frith, C. D., Poline, J.-B., Heather, J. D., and Frackowiak, R. S. J. 1995. Spatial registration and normalization of images. *Hum. Brain Mapp.* **2:** 165–189.

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., and Turner, R. 1996. Movement-related effects in fMRI time-series. *Magn. Reson. Med.* **35:** 346–355.

Frouin, V., Messegue, E., and Mangin, J.-F. 1997. Assessment of two fMRI motion correction algorithms. *Hum. Brain Mapp.* **5:** S458.

Grootoonk, S., Hutton, C., Ashburner, J., Howseman, A. M., Josephs, O., Rees, G., Friston, K. J., and Turner, R. 2000. Characterization and correction of interpolation effects in the realignment of fMRI time series. *NeuroImage* **11:** 49–57.

Hajnal, J. V., Myers, R., Oatridge, A., Schwieso, J. E., Young, I. R., and Bydder, G. M. 1994. Artifacts due to stimulus correlated motion in functional imaging of the brain. *Magn. Reson. Med.* **31:** 283–291.

Hill, D. L. G., Batchelor, P. G., Holden, M., and Hawkes, D. J., 2001. Medical image registration. *Phys. Med. Biol.* **46:** R1–R45.

Holden, M., Hill, D. L. G., Denton, E. R. E., Jarosz, J. M., Cox, T. C. S., Rohlfing, T., Goodey, J., and Hawkes, D. J. 2000. Voxel similarity measures for 3D serial MR brain image registration. *IEEE Trans. Med. Imag.* **19:** 94–102.

Jezzard, P., and Clare, S., 1999. Sources of distortions in functional MRI data. *Hum. Brain Mapp.* **8:** 80–85.

Jiang, A. P., Kennedy, D. N., Baker, J. R., Weisskoff, R. M., Tootell, R. B. H., Woods, R. P., Benson, R. R., Kwong, K. K., Brady, T. J., Rosen, B. R., and Belliveau, J. W. 1995. Motion detection and correction in functional MR imaging. *Hum. Brain Mapp.* **3:** 224–235.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. 1997. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imag.* **16:** 187–198.

Maintz, J. B. A., and Viergever, M. A. 1998. A survey of medical image registration. *Med. Image Anal.* **2:** 1–36.

Meer, P., Mintz, D., and Rosenfeld, A., 1991. Robust regression methods for computer vision: A review. *Int. J. Comput. Vision* **6:** 59–70.

Meyer, C. R., Boes, J. L., Kim, B., Bland, P. H., Zasadny, K. R., Kison, P. V., Koral, K., Frey, K. A., and Wahl, R. L. 1997. Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. *Med. Image Anal.* **1:** 195–206.

Nikou, C., Heitz, F., Armspach, J.-P., Namer, I.-J., and Grucker, D. 1998. Registration of MR/MR and MR/SPECT brain images by fast stochastic optimization of robust voxel similarity measures. *NeuroImage* **8:** 30–43.

Pluim, J. P. W., Maintz, J. B., and Viergever, M. 2000a. Interpolation artifacts in mutual information-based image registration. *Comput. Vision Image Understand.* **77:** 211–232.

Pluim, J. P. W., Maintz, J. B., and Viergever, M. 2000b. Image registration by maximization of combined mutual information and gradient information. In *Proc. MICCAI'00, Pittsburgh, USA, LNCS-1935,* pp. 452–461. Springer-Verlag, Berlin.

Robson, M. D., Gatenby, J. C., Anderson, A. W., and Gore, J. C. 1997. Practical considerations when correcting for movement-related effects present in fMRI time-series. In *Proc. ISMRM 5th Annual Meeting, Vancouver,* p. 1681.

Roche, A., Malandain, G., Pennec, X., and Ayache, N. 1998. The correlation ratio as a new similarity measure for multimodal image registration. In *Proc. MICCAI'98, Cambridge, USA, LNCS-1496,* pp. 1115–1124. Springer-Verlag, Berlin.

Roche, A., Pennec, X., Rudolph, M., Auer, D. P., Malandain, G., Ourselin, S., Auer, L. M., and Ayache, N. 2000. Generalized correlation ratio for rigid registration of 3D ultrasound with MR images. In *Proc. MICCAI'00, Pittsburgh, USA, LNCS-1935,* pp. 567–577. Springer-Verlag, Berlin.

Studholme, C., Hill, D. L. G., and Hawkes, D. J. 1997. Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. *Med. Phys.* **24:** 25–35.

Studholme, C., Hill, D. L. G., and Hawkes, D. J. 1999. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognit.* **32:** 71–86.

Turner, R., Jezzard, P., Wen, H., Kwong, K. K., Le Bihan, D., Zeffiro, T., and 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.

Unser, M., Aldroubi, A., and Eden, M. 1993a. B-spline signal processing. Part I. Theory. *IEEE Trans. Signal Proc.* **41:** 821–833.

Unser, N., Aldroubi, A., and Eden, M. 1993b. B-spline signal processing. Part II. Efficient design and applications. *IEEE Trans. Signal Proc.* **41:** 834–848.

Viola, P., and Wells, W. M. 1997. Alignment by maximization of mutual information. *Int. J. Comput. Vision* **24:** 137–154.

Wells, W. M., Viola, P., Atsumi, H., and Nakajima, S. 1996. Multimodal volume registration by maximization of mutual information. *Med. Image Anal.* **1:** 35–51.

West, J., *et al.* 1997. Comparison and evaluation of retrospective intermodality brain image registration techniques. *J. Comput. Assisted Tomogr.* **21:** 554–566.

Woods, R. P., Cherry, S. R., and Mazziotta, J. C. 1992. Rapid automated algorithm for aligning and reslicing PET images. *J. Comput. Assisted Tomogr.* **16:** 620–633.

Woods, R. P., Grafton, S. T., Holmes, C. J., Cherry, S. R., and Mazziotta, J. C. 1998. Automated image registration. I. General methods and intrasubject, intramodality validation. *J. Comput. Assisted Tomogr.* **22:** 139–152.

Wu, D. H., Lewin, J. S., and Duerk, J. L. 1997. Inadequacy of motion correction algorithms in functional MRI: Role of susceptibility-induced artifacts. *J. Magn. Reson. Image* **7:** 365–370.