Changes in Effective Connectivity Models in the Presence of Task-Related Motion: An fMRI Study

Maria Gavrilescu,^{1,2,5*} Geoffrey W. Stuart,^{3,5} Anthony Waites,⁴
Graeme Jackson,⁴ Imants D. Svalbe,² and Gary F. Egan^{1,5}

¹Howard Florey Institute, University of Melbourne, Melbourne, Australia

²School of Physics and Materials Engineering, Monash University, Melbourne, Australia

³Mental Health Research Institute of Victoria, Melbourne, Australia

⁴Brain Research Institute, Austin Repatriation Center, Melbourne, Australia

⁵Centre for Neuroscience, University of Melbourne, Melbourne, Australia

Abstract: We investigated the effects of motion-correction strategy and time course selection method when structural equation modeling is applied to fMRI data in the presence of task-related motion. Three motion-correction methods were employed for a group of 12 subjects performing an orthographic lexical retrieval task: (1) a rigid body realignment as implemented in SPM99, (2) a rigid body realignment combined with the inclusion of motion parameters in the statistical model, and (3) the FLIRT motion correction followed by an ICA analysis aiming to identify and remove the motion-related components and the ghosting artifacts. For each motion correction, the time courses of the activated regions were selected in three ways: (1) using the voxels with the highest Z scores, (2) using the average across all the statistically significant voxels in the region of interest, and (3) using a within-region, across-subjects, singular value decomposition. The resulting models of effective connectivity were markedly different, although the activation pattern was not substantially altered by the motion-correction method. Higher values for the path coefficients were obtained for the models fitted to the covariance matrices based on the average time courses than for the covariance matrices based on a single voxel time course. Our results suggest caution with the interpretation of task-induced changes in effective connectivity since, for higher-order cognitive brain functions, multiple models can be fitted to a given data set and these models cannot be rejected on an anatomical or cognitive basis. *Hum. Brain Mapp.* 21:49–63, 2004. © 2003 Wiley-Liss, Inc.

Key words: structural equation modeling; path analysis; multivariate normality; functional connectivity; effective connectivity

INTRODUCTION

Functional neuroimaging has been extensively used in the last decade to identify segregated brain areas activated by sensorimotor or cognitive tasks. As brain functions are a result of functional integration and complex interactions between remote brain regions [Horwitz, 1994], new concepts and analysis tools have been introduced to study these interactions. The study of interactions between brain areas is based on two major concepts: *functional connectivity*, defined as “temporal correlation between spatially separated neurophysiological measurements” [Friston et al., 1993a], and *effective connectivity* is defined as “the influence one neuronal system exerts over another” [Friston et al., 1993b]. Functional connectivity reveals common patterns in temporal

Contract grant sponsor: Neuroinformatics Platform, Neurosciences Victoria; Contract grant sponsor: Monash University, Melbourne; Contract grant sponsor: NARSAD, New York.

*Correspondence to: Maria Gavrilescu, Howard Florey Institute, University of Melbourne, 3010 Victoria, Australia.
E-mail: maria@pcomm.hfi.unimelb.edu.au

Received for publication 14 November 2002; accepted 30 September 2003

DOI 10.1002/hbm.10149

fluctuations of remote brain areas while effective connectivity goes one step further making assumptions about the structure and directionality of the interactions. A central assumption underlying these definitions is that a similarity in the pattern of the haemodynamic response across a temporal sequence of fMRI scans implies some underlying connectivity between regions of the brain. This “connectivity signal” must be observable at the temporal limit determined by the scan rate, which is quite slow in relation to most neural processes. For fMRI data, functional connectivity can be estimated by using either resting state data or data sets acquired during an activation paradigm. Recent studies have suggested that the source of connectivity in fMRI may reside in very low-frequency common fluctuations (<0.08 Hz) between distant brain areas [Hampson et al., 2002; Lowe et al., 1998; Xiong et al. 1999]. According to these studies, functional connectivity can be revealed in the resting state and is due to the spontaneous firing of interconnected neurons. Estimating the connectivity in the low-frequency domain is not confounded by the common input from an external stimulus; however, it is complicated by the $1/f$ -like nature of the fMRI noise [Bullmore et al., 2001]. Conversely, some other studies have suggested that meaningful correlations can be observed in activation paradigms at frequencies higher than that of the task [Honey et al., 2002; Rowe et al., 2002]. In this study, we are interested in functional connectivity as estimated in activation studies. An important stream of connectivity studies was generated by the introduction of psychophysiological interactions [Friston et al., 1997] for factorial experimental designs. These studies are focused on detecting context-sensitive changes in connectivity. We are not concerned here with this type of application but rather with the estimation of functional connectivity in relation to a subsequent effective connectivity analysis via structural equation modeling (SEM).

Two issues need to be clarified in relation to the above definition of functional connectivity when estimated for activation paradigms. First, as Friston [1994] has pointed out, when two brain regions respond in a similar manner to an external stimulus, this does not necessarily imply a functional connection. It is accepted in many studies that when the scan series is derived from an activation paradigm, the common input from the stimulus should be removed before the temporal correlations are estimated. To do this, one can either estimate the temporal correlations within condition [Gonçalves et al., 2001; Honey et al., 2002; McIntosh and Gonzales-Lima, 1994] or the stimulus can be regressed out and the correlations estimated on residuals [Rowe et al., 2002]. Second, imaging artifacts (signal dropout, drift, ghosting) as well as the physiological noise (cardiac and respiratory) can contaminate the temporal correlations between regions [Krüger and Glover, 2002; Lund, 2001; Peltier and Noll, 2002]. These effects need to be identified and eliminated from the data prior to correlation estimation since they are obviously not related to functional connectivity.

For statistical analyses based on the general linear model, this contamination may be particularly problematic in the

presence of confounds (variables of no interest significantly correlated with the experimental paradigm). The preprocessing strategies applied to the data may, in this context, influence the way that the shared variance is divided between the task-related regressor and the confounds. This, in turn, may affect the functional connectivity estimates. An important source of confounds for fMRI data is the task-correlated motion (motion parameters are significantly correlated with the experimental paradigm).

We estimated the functional connectivity for an orthographic lexical retrieval (OLR) fMRI data set in the presence of task-correlated motion. In this context, we investigated how varying the motion-correction strategies may affect the temporal correlation estimates. Since no general consensus has been reached regarding how to deal with task-correlated motion, we have employed three motion-correction strategies currently used in fMRI studies: (1) a rigid body realignment as implemented in SPM99; (2) a rigid body realignment with inclusion of the motion parameters in the design matrix as user defined regressors; and (3) the FLIRT motion-correction algorithm followed by an independent component analysis (ICA) to removed motion-correlated components and other artifacts. With significant task-correlated motion, the first method is likely to incorrectly distribute motion-induced variance to the task-related regressor, while the second method is expected to share part of the task-induced variance with the motion parameters. For the third method, after the FLIRT realignment, a spatial independent components analysis (ICA) was performed for each subject in order to decompose the full 4-D mean-centred and variance-normalised data set into spatial statistically independent component maps and associated time courses [McKeown et al., 1998]. The time courses of two spatially independent components can be correlated in any degree. Under conditions of task-correlated motion, this property can be used to separate the task and the motion-induced variance when the two sources of variance are expressed in separate spatial maps. The last method is also useful in order to identify imaging artifacts such as ghosting and signal dropout that cannot be identified with more conventional preprocessing techniques.

Assuming that problems in estimating the correlations that represent functional connectivity can be resolved, it may then be possible to carry out an analysis of effective connectivity. The most common method used to analyse effective connectivity is structural equation modeling (SEM). SEM is a general analytic framework for the analysis of covariance matrices. The specific form of analysis used to study effective connectivity is an extension of the traditional linear regression technique of path analysis [Wright, 1934]. The use of SEM for functional neuroimaging data was introduced by McIntosh and Gonzales-Lima [1994] and has subsequently been applied to a wide range of functional MRI studies of attention [Büchel and Friston, 1997], learning [Büchel et al., 1999] memory [Maguire et al., 2000], auditory processing [Gonçalves et al., 2001], and working memory [Honey et al., 2002]. SEM can be used to assess anatomically

plausible models involving direct and indirect connections between brain regions. For a given model, SEM estimates a set of free parameters (connection weights and residual variances) that produces a covariance matrix as close as possible to the sample covariance matrix. A statistical inference about the model is made based on a discrepancy measure (e.g., maximum likelihood function) between the covariance matrix implied by the model and the sample covariance matrix. The discrepancy measure, multiplied by $N-1$ (where N is the number of independent observations), is χ^2 distributed when the model is correct and the observations are multivariate normally distributed [for details, see Hoyle, 1995; McIntosh and Gonzales-Lima, 1994]. Some consideration of the effects of violation of these two major assumptions is warranted. In addition, some adjustment needs to be made to deal with the fact that sequential scans are not statistically independent of each other.

The discrepancy measure asymptotically approaches a χ^2 distribution only for multivariate normally distributed variables. When the multivariate normality assumption is violated, simulation studies [Chou and Bentler, 1995] demonstrated that regardless of the discrepancy measure, the χ^2 value is overestimated and the standard errors of the parameters are underestimated [MacCallum, 1995; West et al., 1995]. In other words, the rate of rejected models is bigger than the nominal type I error rate, and path coefficients that are not significantly different from zero at the population level can become significant for that particular sample. However, the effects of deviation from multivariate normality depend on the sample size and model complexity [Chou and Bentler, 1995; Raykov and Widaman, 1995]. Raykov and Widaman [1995] indicated that, depending on the complexity of the model and the extent of deviation from multivariate normality, between 5 and 20 observations per parameter (path coefficients and residual variances) may be required. Thus, an assessment of multivariate normality is essential for a proper interpretation of SEM results.

The other major assumption regards the correctness of the model. For functional neuroimaging data, this problem has two aspects. First, all necessary and sufficient brain regions involved in a certain network have to be identified. Second, the proper set of connections (including the direction of influence) has to be chosen. Region identification usually relies on statistical analyses of the data set using univariate methods [Büchel et al., 1997; Gonçalves et al., 2001] or multivariate methods [Della-Maggiore et al., 2000; Jennings et al., 1998; McIntosh et al., 1999; Nyberg et al., 1996]. However, some regions involved in the task performance may not be identified as activated due to low sensitivity, weak signal changes, and also because of a limited field of view (e.g., subcortical structures such as thalamus). Confounding effects such as attention modulation and learning may produce activated regions that are not specific to the function of interest. Price and Friston [1999] recommended the use of patients with specific lesions to identify the necessary and sufficient network for one particular brain function. Similarly, McIntosh and Gonzales-Lima [1994] suggested that

SEM models can be objectively constrained by using neuro-anatomical information about the connections between brain areas included in the models. However, this information is mainly based on primate studies and cannot always be easily extended to the human brain [Bullmore et al., 2000]. There are brain regions where the information about anatomical connections is incomplete [e.g., auditory system, Gonçalves et al., 2001] and also there are brain regions with a large number of direct and reciprocal connections, where a complete model is impossible to estimate [Bullmore et al., 2000]. In such a case, only a subset of possible connections is chosen for analysis [Bullmore et al., 2000; McIntosh, 1999]. Although this choice is based on physiological and cognitive theories about the task under study, this selection can be somewhat arbitrary. For example, in a network with 4 regions there are 12 possible connections. When none of these connections can be eliminated on an anatomical basis, by fixing the values of the residual variances (that proportion of the variance in each region that is treated as noise and is not explained by the model), any model with up to 9 connections can be estimated. A total of 4,016 models¹ fall in this category and it is expected that at least some of them will fit the covariance matrix. Some of these models can be rejected based on cognitive theories (e.g., primary auditory regions are assumed to drive language areas). It is, however, very likely that a reasonably large number of models will survive the screening (mainly due to reciprocal connections), especially for higher order cognitive functions such as memory or attention. In general, the study of effective connectivity with SEM is complicated by the non-uniqueness of the model under conditions of weak anatomical constraints. There are two potential sources of variability that can lead to multiple models fitting the same data set. The first one pertains to the SEM itself since for any given covariance matrix, in principle, a large number of good fitting models can be constructed that will differ only in the substantive interpretation [MacCallum et al., 1993]. The second source of variability resides in the complex spatio-temporal structure of fMRI data that can lead to different identification of functional connectivity with different preprocessing strategies and/or time course selection methods. These differences can be enhanced in the presence of confounds. We investigated here only the second source of variability.

In this study, SEM was employed to assess the stability of functional and effective connectivity estimates and the validity of SEM assumptions for an OLR fMRI data set in the presence of task-related motion. Of particular interest were the effects of different pre-processing strategies used to remove artifacts due to motion, as well as other scanning artifacts. In order to remove covariance due to task-corre-

¹For a model with $q = 4$ regions, $1/2q(q+1) = 10$ connections (the number of independent elements in the covariance matrix) can be estimated. However, in order to assess the model fit at least one degree of freedom is necessary; therefore, we considered here only the models with up to 9 connections. The total number of models can be then calculated as $\sum_{i=1}^9 C_{12}^i$.

lated motion, we employed the three methods described above. In the present study, we did not consider the effects of cardiac and respiratory noise since these factors were not measured in our sample data set. For each preprocessing strategy, we have also employed the most common methods for time course selection: using the voxel with the highest Z score for each region [Della Maggiore et al., 2000; Gonçalves et al., 2001; Jennings et al., 1998; Maguire et al., 2000; McIntosh and Gonzales-Lima 1994; Nyberg et al., 1996], using the average time course across selected voxels in each region [Büchel et al., 1999; Büchel and Friston, 1997; Fletcher et al., 1999; Honey et al., 2002], and using the singular value decomposition (SVD) procedure proposed by Bullmore et al. [2000]. To our knowledge, these methods have not been previously compared for a given data set. A multivariate normality assessment was performed for each sample. Due to a lack of strong anatomical constraints, initial models were constructed in a stepwise manner [Bullmore et al., 2000; Honey et al., 2002]. Cross-validation was also used to determine whether different effective connectivity models were due to small chance variations in covariance matrices or arise from substantial differences due to the pre-processing strategies.

SUBJECTS AND METHODS

Preprocessing and Statistical Analysis

Data description

Twelve healthy volunteers were scanned using a 3T G.E. Signa LX whole body scanner while performing an orthographic lexical retrieval (OLR) task. During the task period, the subjects were asked to silently generate as many words as possible starting with a given letter. Two letters were presented on a screen in the active condition (one letter every 18 sec). The task was presented as a block design paradigm with a period of 72 sec. In the rest condition, the subjects were asked to fixate on a crosshair at the centre of the screen. The two conditions were matched in terms of luminosity and contrast. A total of 94 full brain T_2^* weighted images were acquired for each subject with a TR of 3.6 sec, FA = 40, 2×2 mm in plane resolution and 22 trans-axial slices of 4-mm thickness and 1-mm gap. The first four scans were discarded for each subject to allow for T1 saturation effects. The stimulus presentation started and ended with a rest condition providing four complete trials for each subject.

Motion correction strategies

Three motion-correction strategies were employed in order to deal with the task-correlated motion identified in this study. For the first two methods, the images for each subject were aligned to the fifth image in the series using a rigid body 6 df realignment as implemented in SPM99 (online at <http://www.fil.ion.ucl.ac.uk/spm>). The optimal transformation in SPM99 is found by minimising the sum of the squared differences between two images. A 3-D Gaussian smoothing (FWHM = 8 mm) is applied before the realign-

ment. For the first method (spm), the realigned images were spatially normalised and the statistical analysis was performed in spm99. For the second motion-correction method (spm_mp), the motion parameters (3 rotations and 3 translations) were included as user defined regressors in the design matrix. The correlation coefficient (r) between the six motion parameters and the experimental paradigm was estimated for each subject, with a value $r > 0.3$ representing a significant correlation.

The third motion-correction method employed FLIRT realignment software (online <http://www.fmrib.ox.ac.uk/fsl>). By default, all images are realigned to the middle image in the time courses (the 45th image for this study). The cost function for the realignment algorithm is the normalised correlation and the optimisation is based on a multi resolution approach [Jenkinson et al., 2002]. After realignment, a spatial ICA decomposition as implemented in Melodic v. 1.1 (online at <http://www.fmrib.ox.ac.uk/fsl/melodic>) was applied for each subject. In this implementation, ICA is a fully automatic algorithm, which calculates the number of independent components using a probabilistic principal components analysis (PPCA) approach, combined with a Bayesian model order selection [Beckmann and Smith, 2002]. The task-related component was identified by ranking the components after the correlation with the experimental paradigm [Esposito et al., 2002]. The correlation coefficient with the motion parameters was estimated for each independent component and a correlation $r > 0.3$ was considered significant. By visual inspection of the temporal courses and spatial patterns, imaging artifact components were also identified. The components that were significantly correlated with the motion parameters ($r > 0.3$) and the artifact components were then removed from the data by projecting the data set onto the remaining components. The resulting images were then spatially normalised and preprocessed as for the first two methods.

Statistical analysis

All statistical analyses were performed using the general linear model implementation in SPM99. After motion correction, the mean functional image was spatially normalised to the EPI template using the nonlinear algorithm based on discrete cosine deformation functions ($7 \times 8 \times 7$ functions). The functional images were resliced using bilinear interpolation. A first level analysis was performed for each subject after smoothing with a 10-mm FWHM isotropic Gaussian kernel, global normalisation by proportional scaling and high pass filtering (151 sec). The activated regions were identified via a random effects analysis at the group level using a false discovery rate (fdr) threshold $fdr < 0.05$ [Genovese et al., 2002].

Functional Connectivity Estimation: Region Identification and Time Course Selection

The brain regions that were candidates for the investigated network were chosen from the activation maps, based on currently accepted theories about language processing.

For each motion-correction algorithm, the covariance matrix for the selected regions was estimated based on three methods of time course selection: (1) using the time courses for the voxels with the highest Z score for each region (single voxel time courses), (2) using the average across all the voxels exceeding a false discovery rate of 0.05 in a 10-mm radius sphere around the voxel with the highest Z score for each region (average time course), and (3) using a SVD to identify the task related component for each region (SVD time courses) [Bullmore et al., 2000]. This approach is based on the SVD for a $M \times NS$ matrix constructed for each region with one row per subject (M subjects) and one column per scan (NS scans/subject). The eigenvectors or “eigentimeseries” [Bullmore et al., 2000] are ranked according to the amount of variance explained. The task-related component was identified as the component presenting the highest correlation with the experimental paradigm. These components (one $1 \times NS$ time course/region) were then used to estimate the covariance matrix. The SVD method was applied for both the single voxel and the average time courses. All time courses were normalised (zero mean and unit standard deviation) prior to covariance estimation.

To discard the common input from the external stimulus for the single voxel and average time courses, the covariance matrices were estimated within condition and the time courses were concatenated across subjects (480 observations for the active condition and 600 observations for the baseline). The number of observations was subsequently corrected for temporal autocorrelation using the effective number of degrees of freedom [Worsley and Friston, 1995]. The covariance matrices based on the SVD time series were estimated across conditions due to concerns regarding the sample size (90 observations/eigentimeseries) for the subsequent effective connectivity analysis (see Model Construction and Estimation). The common stimulus input was, therefore, not discarded from these time courses.

Effective Connectivity Estimation

Residual variances

For functional neuroimaging data, often the number of plausible connections exceeds the number of degrees of freedom. Therefore, a common practice is to fix the values of the residual variances. McIntosh and Gonzales-Lima [1994] suggested arbitrary values ranging from 0.35 to 0.50 of the variance depending on the number of inputs and outputs for a given region: the bigger the number of inputs, the smaller the residual variance. In practice, some authors have arbitrarily fixed the residual variances [Della Maggiore et al., 2000; Gonçalves et al., 2001; McIntosh and Gonzales-Lima 1994, 1999b; Nyberg et al., 1996], while others have estimated them as free parameters in the model [Büchel et al., 1999; Fletcher et al., 1999]. In this study, we adopted the estimation of residual variances based on objective measures as suggested by Bullmore et al. [2000]. Following the SVD for the matrix constructed for the time course of each region (single voxel and average time courses) with one row per

subject and one column per scan, the residual variances (Ψ_i) can be estimated by the ratio of the eigenvalue corresponding to the task related eigentimeseries (λ_k) to the sum of all eigenvalues (λ_j , $j = 1, \dots, m$):

$$\Psi_i = 1 - \frac{\lambda_k^2}{\sum_{j=1}^m \lambda_j^2} \quad (1)$$

This method was applied to estimate the residual variances for all time course selection methods.

Normality assessment

We assessed the normality of the time courses distributions both at the univariate and multivariate level. The univariate normality was assessed by measuring the skewness (s , third order moment) and kurtosis (k , the fourth order moment) for each variable and the associated standard errors. The standard error of the skewness can be calculated as $SE_s = (6/N)^{1/2}$ and for kurtosis as $SE_k = (24/N)^{1/2}$, where N is the sample size [Kendall, 1994]. A normal distribution is characterised by a zero skewness and a kurtosis $k = 3$.² A univariate distribution is significantly non-normal when the difference between the skewness and/or kurtosis of the given distribution and the corresponding values for the normal distribution are larger than two standard errors [Kendall, 1994]. Univariate normality does not imply normality for the joint distribution, therefore for SEM it is important to estimate multivariate measures of normality. The multivariate normality was assessed by measuring the multivariate kurtosis as defined by Mardia [1970]. A multivariate kurtosis $k = 0$ indicates a multivariate normal distribution. The multivariate kurtosis is statistically significantly different from zero when the critical ratio (CR, the ratio of multivariate kurtosis to its standard error) is $CR > 1.96$ or $CR < -1.96$. The estimate of skewness and kurtosis was performed using AMOS 4.01 (online at <http://www.smallwaters.com>).

Model construction and estimation

The effective connectivity models were estimated using AMOS v. 4.01 with the maximum likelihood function as the discrepancy measure between the sample and the implied covariance matrices. The investigated brain network is a typical case of network with weak anatomical constraints. Therefore, rather than testing a theoretical model, we constructed the models using the step-up procedure, based on modification indices, as described in Bullmore et al. [2000] and applied by Honey et al. [2002]. The process starts with a model where all the path coefficients are set to zero. At

²Some computer packages subtract 3 from the estimated kurtosis so that a normal distribution is characterised by a zero kurtosis. This is the convention followed in our study to report the kurtosis values.

each step, the path coefficient with the maximum modification index is set free to be evaluated. Modification indices or Lagrangian multipliers [Bullmore et al., 2000] are functions of the derivatives of the maximum likelihood function in respect to all parameters (free and constrained). They indicate the change in the χ^2 value by setting free a parameter that was initially constrained [Hoyle, 1995]. When a model fitting the data was found or all the modification indices were zero, the process was stopped. For the resulting model, we have investigated the significance of the path coefficients using the critical ratio (CR, the ratio of path coefficients to their standard errors). CR has a Z distribution [Hoyle, 1995] and the estimated path coefficients with $CR > 1.96$ or $CR < -1.96$ are significantly different from zero. With deviations from multivariate normality, a higher CR threshold ($CR > 3$ or $CR < -3$) was suggested [Hoyle, 1995]. The path coefficients with $-3 < CR < 3$ were removed one by one and the resulting models were retested. The models with a good fit ($P > 0.05$) and the smallest number of nonsignificant path coefficients were then reported. The significance of the path coefficients was assessed using the critical ratio.

As Chou and Bentler [1995] suggested, for samples with less than 100 observations, the large variability of the estimates makes it impossible the generalisation of the results to the population level. Therefore, in this study we did not perform within subject SEM estimates due to the small samples (90 scans per subject, 40 scans during task, and 50 scans during baseline condition).

After we found a model that fitted each covariance matrix, we cross-tested the fit of these models across all covariance matrices estimated for each of the three motion correction strategies and the three time course selection methods.

Condition effects

A study of the condition effects was performed using the stacked models approach [Hoyle, 1995; McIntosh and Gonzales-Lima, 1994]. The data matrix was separated in task time courses (480 observations at the group level) and baseline time courses (600 observations at the group level). The models were constructed for the task and baseline time courses [Büchel and Friston, 1997]. Each path coefficient was then tested for task effects by using the χ^2_{diff} between the null model (n degrees of freedom) where all coefficients were constrained to be equal across conditions, with the alternative model ($n+1$ degrees of freedom) where one parameter at a time was allowed to be different. The change in the χ^2 value produced by $\chi^2_{null} - \chi^2_{alternative}$ with one degree of freedom that was then used to assess the null hypothesis that a single estimate of the path coefficient is sufficient for both conditions. A probability value $P < 0.05$ indicates a significant difference in the path coefficients across conditions.

RESULTS

Preprocessing and Statistical Analysis

The amplitude of the motion as detected via SPM99 and FLIRT was small, less than 2 mm translation and less than 2

TABLE I. Correlation coefficients between the experimental paradigm and the motion parameters as estimated via SPM99

Subject no.	x_trans	y_trans	z_trans	x_rot	y_rot	z_rot
1	-.64	-.02	.09	-.01	.37	-.15
2	-.08	-.67	-.23	.48	.59	-.10
3	.12	-.30	.08	-.10	-.05	.05
4	-.23	-.03	-.10	.10	.14	-.39
5	-.06	-.56	.22	.03	.36	.06
6	.02	.21	-.33	.02	.01	-.14
7	-.08	-.50	.23	-.30	.14	.04
8	-.05	-.19	-.16	.24	.17	-.07
9	-.03	.05	.06	.04	.05	-.07
10	-.47	-.40	-.13	-.03	-.38	-.64
11	-.12	.10	-.02	.17	-.11	-.35
12	.18	-.17	.10	-.08	-.16	.02

* trans = translation; rot = rotation.

degrees rotation across subjects. However, both realignment algorithms revealed significant correlation ($r > 0.3$) between at least one motion parameter and the experimental paradigm for 8 out of 12 subjects. The correlation coefficients for the SPM99 motion parameter estimates are presented in Table I.

The same five statistically significant activated regions ($fdr < 0.05$) were identified with the random effects analysis for each of the three preprocessing strategies (Fig. 1A, Table II): left and right inferior frontal gyrus (LIFG and RIFG), left medial frontal gyrus (LMFG), anterior cingulate (AC), and angular gyrus (AG). The posterior cingulate (PC) was significantly deactivated. This deactivation was not an artifact due to global normalisation since it was statistically significant with both proportional scaling and grand mean scaling normalisation [Gavrilesco et al., 2002].

The FLIRT motion correction method induced slightly higher Z scores and a slightly higher number of voxels in the regions of interest compared with spm, which in turn slightly outperformed spm_mp. The position of the voxel with the highest Z score for each region varied across motion correction methods. The largest variation was for the PC in the z coordinate (18 mm) while for the other regions the differences were typically around 6 mm for all coordinates.

The spatial pattern of activation was consistent both for the first and second level analyses. For spm and FLIRT motion correction, 11 of 12 subjects showed activation in all regions identified at the group level. The activation in RIFG was not statistically significant for the remaining subject, but BOLD signal increase was evident at a lower probability threshold ($P_{uncorrected} < 0.001$). With spm_mp, for 6 of 12 subjects all regions were observed ($P_{uncorrected} < 0.05$) with different regions undetected across the remaining 6 subjects. However, all regions were identified for these subjects at a lower probability threshold ($P_{uncorrected} < 0.001$). All these 6 subjects presented a significant correlation ($r > 0.3$) between the experimental paradigm and at least one of the motion parameters.

Figure 1.

A: The regions activated by the OLR task: LIFG, left inferior frontal gyrus; RIFG, right inferior frontal gyrus; LMFG, left middle frontal gyrus; AC, anterior cingulate; AG, angular gyrus; PC, posterior cingulate. PC was significantly deactivated. The activations were obtained in a random effects analysis ($\text{fdr} < 0.05$) for a group of 12 subjects with the spm motion correction algorithm. **B:** Examples of spatial independent components and associated time courses: task related component ($r = 0.77$); **C:** motion-related

component that is also significantly correlated with the task ($r = 0.80$); **D:** drift-related component ($r = 0.37$); **E:** component expressing a combination of ghosting artifacts and motion by susceptibility interaction ($r = 0.06$). r represents the correlation coefficient between the experimental paradigm and the time course of each component. Components such as C–E were regressed out from the data. The ICA images are in radiological orientation.

COLO

TABLE II. Talairach Coordinates of the voxels with the highest Z score for each region

Region	Talairach coordinates								
	spm			spm_mp			flirt		
	x	y	z	x	y	z	x	y	z
Left inferior frontal gyrus (LIFG)	-42	22	-10	-42	16	-10	-42	22	-10
Right inferior frontal gyrus (RIFG)	42	22	-8	36	22	0	42	22	-8
Left medial frontal gyrus (LMFG)	-52	12	22	-50	8	22	-52	12	22
Angular gyrus (AG)	-44	-48	40	-44	-50	42	-44	-48	40
Anterior cyngulate (AC)	-2	6	54	2	2	58	-2	6	54
Posterior cyngulate (PC)	2	-52	38	2	-66	38	2	-52	38

ICA decomposition provided 26 to 54 independent components across subjects. For each subject, the time course for only one component was strongly correlated with the task ($0.57 < r < 0.89$). The spatial pattern associated with that component was similar to the activation pattern as revealed by the random effects analysis (Fig. 1B). For the components significantly correlated ($r > 0.3$) with at least one motion parameter, visual inspection showed that they were associated with signal drift over time, spatial patterns around the edges of the brain or in the central ventricles. All these components were regressed from the data. For some subjects, independent components expressing a combination of Nyquist ghosting artifacts and motion by susceptibility interaction were identified. Although not correlated with the motion parameters, these components were also removed since they were obviously not connectivity related. Some examples of removed components are presented in Fig. 1C–E).

Functional Connectivity Estimation. Region identification and Time Course Selection

All significantly activated regions identified in this study were consistent with cognitive models of language processing. The major differences in functional connectivity estimation across the time course selection and motion correction methods are summarised in Figure 2. For simplicity, we have not included the correlation coefficients for AG (see Model Assessment). The motion-correction strategy has a major influence on the correlation matrix determined via SVD time courses (Fig. 2A). When SVD was used in combination with FLIRT and ICA, the resulting correlation matrix was uniform (all elements ~ 0.9) while spm_mp gave very small values for all interregional correlations. Possible explanations for this striking effect will be addressed in the Discussion. The identification of the task-related eigentime-series was also influenced by the motion correction method. For spm and FLIRT motion correction, the first eigentime-series had the highest correlation with the paradigm for all but one region (RIFG for spm and LIFG for FLIRT) where the task-related eigentime-series was ranked second. However, for spm_mp motion correction, the rank of the task-related eigentime-series ranged from 1 to 4 across regions.

For the average and single voxel time courses, only the correlation matrices corresponding to the active condition are presented in Figure 2B for the average and Figure 2C for the single voxel time courses. A similar trend was observed for the baseline time courses. These matrices were less affected by the preprocessing strategy. As a general trend, the largest correlation coefficients were obtained with FLIRT, while spm induced slightly higher correlation values than spm_mp. It can be noted also that the correlation coefficients obtained for the average time courses were, in general, bigger than those obtained for the single voxel time courses. This trend is reversed for the covariance matrices where the values for the single voxel time courses were three orders of magnitude larger than the values for the average time courses.

Effective Connectivity Estimation

Residual variances

The estimated residual variances via relation (1) were in general bigger with smp_mp, both for average and single voxel time courses, while FLIRT provided the smallest values (Table III). For each motion-correction method, the residual variances estimated for the average time courses were, in general, similar to those estimated for the single voxel time courses.

Normality assessment

Significant deviations from univariate normality were observed for some regions over all motion correction algorithms and time course selection methods (the estimated skewness and/or kurtosis were larger than two standard errors). Since the univariate normality is a necessary (but not sufficient) condition for the normality of the joint distribution, it is expected that the multivariate normality would be also violated. Indeed, the multivariate distributions were significantly non-normal for all three motion-correction methods when the average time course and the single voxel time courses were used to calculate the covariance matrix (Table IV, models D to G). However, the SVD time courses were, surprisingly, multivariate normally distributed for all motion correction methods ($-1.96 < CR < 1.96$).

Figure 2.

Correlation coefficients between the activated regions estimated using **A**: The across-conditions time courses based on the within region SVD method (see text for details); **B**: the average time courses for all the significantly activated voxels ($\text{fdr} < 0.05$) in a

sphere with 10-mm radius during task condition for each region; and **C**: the time courses of the voxels with the highest Z scores during task condition.

Model assessment

The models fitted to the covariance matrices with all six regions contained a large number of connections. Most of the path coefficients pointing to and from AG were not significantly different from zero. Once the AG time courses were excluded from the covariance matrices, the fitted models were simpler and had a smaller number of non-significant path coefficients. Therefore, we concentrated in studying these models. The use of a subset of the activated regions was used before in order to simplify the resulting models [Büchel and Friston, 1997].

Models were fitted to all covariance matrices estimated via the three time course selection methods, for all motion

correction algorithms. For simplicity, only some of these models are presented in Figure 3. However, a model could not be fitted for the covariance matrix based on the SVD time courses with spm motion correction (Figure 3, model A). In this case, all modification indices were zero before the χ^2 value reached significance ($P > 0.05$). In all cases, the residual variances were fixed to the levels indicated in Table III. With the correction for autocorrelation in the fMRI time courses, the effective sample size became $N = 442$ for task time courses, $N = 552$ for baseline time courses, and $N = 82$ for the SVD time courses. The models fitted for the active condition are presented in Figure 3 (D–G). Similar differences in the models across the motion

TABLE III. Residual variances estimated via relation I across the motion correction methods both for the single voxel time courses (voxel) and for the average time courses (average)

Region	spm		spm_mp		flirt	
	Average	Voxel	Average	Voxel	Average	Voxel
LIFG	58.6	59.8	86.6	79.1	42.3	87.5
RIFG	83.9	83.4	83.4	86.2	63.1	64.3
LMFG	40.5	57.8	57.8	81.8	29.3	30.4
AG	66.6	67.1	78.4	90.3	53.5	50.9
AC	44.1	46.1	76.1	81.5	33.1	40.2
PC	47.9	57.6	96.9	96.2	41.4	94.3

Values are expressed as percentage of the total variance.

correction strategies were observed for the baseline condition. It should be noted that for all the fitted models, the other goodness of fit indices calculated by AMOS 4.01 were also in the appropriate range of values for a well-fitting model.

The motion-correction algorithm had a strong impact on the models fitted for the covariance matrices estimated via SVD (Figure 3, models A, B, and C). These models were quite different with respect to connections and in the values of path coefficients.

In general, for all motion-correction methods, the values of the path coefficients were larger for the average time courses than for the single voxel time courses (Figure 3, models D and E). The methods used for motion correction and time course selection introduced three types of differences in the models fitted for the average and one voxel time courses: in the value of path coefficients, in the number of connections, and in the directionality (Figure 3, models E, F, and G). The value of path coefficients varied across methods with spm_mp inducing smaller values than the other two motion correction methods, consistent with the smaller values of the correlation coefficients. The number of connec-

tions identified with each model was also variable across methods. Some connections that were not present in all models were between the LIFG and the AC, the AC and the PC, the LMFG and the RIFG, and the RIFG and the PC. These connections are present only in some models while for the others they were either zero or not statistically significant. The directionality of connections was affected as well. The connections between LIFG and LMFG, AC and PC, LMFG and PC changed direction across models and sometimes became not significant. Path coefficients not significantly different from zero were identified for many models. The path coefficients with a critical ratio $1.96 < CR < 1.96$ are marked by a double star while those with $-3 < CR < 3$ are marked by one star (Figure 3).

Each model reported in Figure 3 was cross-tested against all covariance matrices. None of these models presented a good fit when tested for a covariance matrix different from the one that originally produced the model. For each covariance matrix, the investigation of modification indices pointed to changes in the model that when followed, produced a model identical to the one originally fitted for that particular sample. We have also tested a consensus model derived from the common features across the models in Figure 3. This model contains projections from LIFG to RIFG, projections from LIFG and LMFG to PC (the directionality of these projection varies across models), projection from LIFG and LMFG to AC, and a connection between AC and PC. The consensus model did not fit any of the covariance matrices.

Condition effects

For single voxel and average time courses, we measured the significant differences for all path coefficients in task vs. baseline. The path coefficients that were significantly different ($P < 0.05$ for χ^2_{diff}) are marked by ellipses in Figure 3. The condition effects were different across methods and none of the significant effects was present in all models. In general, during task, some models demonstrated an increase

TABLE IV. Normality assessment and models fit

Model	Univariate normality		Multivariate normality		Model fit		
	Skewness	Kurtosis	Kurtosis	CR	χ^2	df	P
A	0.06 → 0.26	-1.44 → -0.67	2.05	1.16	31.7	9	0.000
B	-0.42 → 0.11	-0.64 → 2.75	2.75	1.56	14.70	10	0.197
C	0.12 → 0.21	-1.56 → -1.13	1.31	0.74	10.40	10	0.406
D	-0.81 → 0.48	0.95 → 4.23	11.51	16.84	1.97	7	0.961
E	-1.08 → 0.51	0.94 → 4.34	28.42	37.29	12.49	8	0.131
F	-0.23 → 0.48	0.50 → 1.67	13.61	17.83	4.50	7	0.715
G	-0.77 → 0.36	0.56 → 2.17	7.81	10.23	11.34	8	0.183

Time courses identified via SVD with all three motion correction algorithms (A, spm; B, spm_mp; C, FLIRT); D: using the single voxel time courses with spm motion correction; E–G: using the average time courses with all three motion correction algorithms (E, spm; F, spm_mp; G, FLIRT). CR denotes the critical ratio for the multivariate kurtosis (see text for definition). For models A–C, the across conditions time courses were used while for models D–G, the presented results refer to the time courses during task condition. Similar trends were observed for the baseline time courses.

TABLE V. The values of the path coefficients that were significantly different across conditions for models D–G

Path coefficient	Model D		Model E		Model F		Model G	
	Task	Baseline	Task	Baseline	Task	Baseline	Task	Baseline
LMFG to LIFG	.29	.15	.38	.23	—	—	—	—
LMFG to AC	.26	.33	.10*	.31	—	—	.38	.45
LMFG to PC	—	—	-.44	-.35	—	—	-.41	-.30
RIFG to AC	.07*	.21	—	—	—	—	—	—
LIFG to PC	—	—	—	—	—	—	-.27	-.18
AC to PC	—	—	-.06**	-.28	-.08*	-.14	0	-.25

* Path coefficients with $-3 < CR < 3$.** Path coefficients with $-1.96 < CR < 1.96$.

in the negative value of the path coefficient between the LIFG and the PC, the LIFG and the LMFG, and the LMFG and the PC. A decrease during task was observed for the path coefficients between the PC and the AC, the LMFG and the AC, and the LIFG and the AC. These effects were not significant in all models but they were similar when present. In Table V are reported path coefficients values for task and baseline time courses across the models D to G in Figure 3.

DISCUSSION

We have examined the differences in the SEM results induced by varying the preprocessing strategies for statistical analysis of fMRI data. We choose an OLR data set since it is a common task used in clinical environment and activates many aspects of the language system in a reliable and robust manner. Furthermore, the brain network revealed in

Figure 3.

Models fitted for the group level covariance matrices in the activation condition. The method to estimate the covariance matrix and the motion correction method are indicated in the legend for each model. *Path coefficients with $-3 < CR < 3$. **Path coefficients with $-1.96 < CR < 1.96$. For models A–C, the across

conditions time courses were used, while for models D–G the presented results refer to the time courses during task condition. The path coefficients marked by an ellipse were significantly different in task vs. baseline.

this study is an example of a network with weak anatomical constraints where all connections are theoretically possible and the choice of a particular effective connectivity model is highly subjective. In this context, we investigated how the functional and effective connectivity estimates were influenced by differences in the noise removal strategy (preprocessing) and differences in accounting for the common stimulus input (time course selection).

Although we found significant task-correlated head movement for the data set under study, the statistical activation maps were remarkably similar across motion-correction methods both at the first level (within subject) and second level (random effects) of analysis. It should be noted that this result does not necessarily imply the consistency of connectivity estimates. However, it may provide the researcher with a false sense of security for subsequent modelling analyses. As we proved in this study, the functional and effective connectivity estimates are more sensitive to the noise removal strategies at the preprocessing stage than the statistical analysis aiming to identify task-activated brain regions.

Due to concerns regarding the small sample size, we did not perform within-subject connectivity estimation. It is expected that large variances in the response to the task across subjects will further complicate the results at the modeling stage. It is also expected that the within-subject results will be sensitive to differences in pre-processing strategies. Studies of the within subject vs. across subjects effective connectivity have already been performed [Büchel and Friston, 1999; Gonçalves et al., 2001] with some contradictory results. While Büchel and Friston [1997] found that the group model was supported by the individual subject models, in a different study Gonçalves et al. [2001] concluded that there was no reproducible pattern either in the values of the path coefficients or in the condition effects across subjects.

With respect to noise removal, we were interested here only in the effects of the motion-correction algorithm. However, other preprocessing strategies (spatial smoothing, filtering in the frequency domain, etc.) can be expected to influence connectivity estimates. The task-correlated motion was detected with two different algorithms, and thus it is unlikely to be an artifact induced by the realignment process [Freire and Mangin, 2001]. One possible explanation for the occurrence of this effect may be the fact that the appearance of a letter on the screen after a relatively long rest period (36 sec) startled the subject and so induced a slight head movement. We wish to emphasise here that we were not interested in finding the best motion-correction method, but rather in proving that when there is shared variance between the task regressor and the motion parameters, this variance is differently distributed by each algorithm resulting in different identification of signal and noise. The inclusion of the motion parameters or their derivatives has already been used in connection with SEM for fMRI data [Büchel et al., 1999; Rowe et al., 2002]. When there is no significant correlation between the motion parameters and the experimental paradigm, this method should increase the

sensitivity of the statistical analysis by modeling unwanted motion-related variance that is not removed by a simple realignment [Bullmore et al., 1999]. With strong and/or widely spread activations, one can expect to identify task-correlated motion simply because the realignment algorithm is intensity based [Freire and Mangin, 2001]. In such a case, the inclusion of the motion parameters in the design matrix is, obviously, not recommended since it is expected to affect the outcome of the statistical analysis. Interestingly, for the data set under study, the functional and effective connectivity estimated were highly dependent upon the motion-correction method while the statistical activation maps were unchanged.

Data de-noising based on ICA represents a promising alternative to deal with the confounding effects of task-correlated motion. It should be noted, however, that we were here in the fortunate situation where only one component could be clearly identified as task related. From our practical experience with ICA, the task effects are usually divided over many components and it can be quite difficult to identify the effects of interest. Since the selection of components to be removed resides solely with the subjective interpretation of the researcher, ICA de-noising presents the potential danger of removing task-induced variance from the data. On the other hand, compared with the usual fMRI data preprocessing, ICA is able to identify sources of artifacts that are obviously not related to functional connectivity such as the artifacts observed in our study. These observations suggest the need for a more objective classification of the independent components.

We studied here the combined effects of noise removal via the preprocessing strategy and the removal of the common input from the external stimulus by performing within condition connectivity estimates. We have also included for comparison the SVD method for time course selection that by using the across conditions time series does not remove the common stimulus input. The estimation of the correlation matrix (functional connectivity) using the SVD algorithm was extremely sensitive to motion correction (Fig. 2A) affecting both the sign and the amplitude of the correlation coefficients and resulting in large differences in the consequent models (Figure 3, models A–C). The *spm_mp* motion correction provided very small correlation values for the SVD time courses. This result confirms that part of the task-related variance was attributed, as expected, to the motion parameters. This observation is also supported by the lower rank of the task-related eigentimeseries. An interesting result is the uniformity of the covariance matrices of the SVD time courses estimated after FLIRT motion correction combined with ICA (all elements were around ± 0.9). This result is not surprising, taking into account that both ICA and SVD were tuned to identify the time course that is the most similar with the external paradigm and to discard other sources of variance as noise. A uniform correlation matrix indicates that there are no preferred connections and the differences between the path coefficients in the model (Figure 3, model C) can be simply induced by differences in

residual variances. Therefore, the uniformity of the covariance matrix can be interpreted in two ways: either the functional connectivity was, at least in part, wrongfully discarded in the “noise” eigentimeseries or there is little functional connectivity between the activated regions and the major source of covariance is the common input from the external stimulus. To avoid the contamination of the SVD eigentimeseries by the common input from the external stimulus, one can perform SVD within condition [Honey et al., 2002]. In this case it is, however, less clear how to select the eigentimeseries associated with the effect of interest since the first eigentimeseries are not always related to the task under study.

For the within-condition connectivity estimates, we considered the differences between single voxel and average time courses. The motion-correction method had a similar impact for both time course selection methods and affected only the amplitude and not the sign of the correlation values (Fig. 2B,C). However, these small differences were enough in order to generate different effective connectivity models. Interestingly, the correlation values obtained for the average time courses were generally bigger than those obtained for the single voxel time courses for all motion correction methods. However, the covariance values for the single voxel time course were three orders of magnitude bigger. This observation suggests that the spatial average across voxels cancelled part of the noise. Since the position of the voxel with the highest Z score is in general determined by the underlying noise structure and is variable across different preprocessing strategies, the average time courses seem to provide a better way to estimate the covariance matrix for modeling purposes. This effect was reflected at effective connectivity stage by bigger path coefficients obtained for the models based on the average time courses (Figure 3, models D and E). This is an important observation since the majority of SEM studies in neuroimaging are based on the single voxel time courses. One problem for the within-condition connectivity estimation is that the common input from the stimulus is considered a simple boxcar. Therefore, this method is problematic for the fMRI data due to the temporal delay introduced by the haemodynamic response function. A possible solution may be to use a boxcar convolved with the haemodynamic response function [Honey et al., 2002]. However, it has been shown that the shape and latency of the haemodynamic response function are variable across subjects and across brain regions [Christoff et al., 2002]. If these differences are characteristic for a given subject and/or region regardless of the task, a simple convolution of the boxcar with a haemodynamic response function that remains constant over subjects and regions may be too simplistic to account for the common stimulus input.

As underlined in Results, the estimated effective connectivity models were different across motion-correction algorithms and time course selection method in the number of connections, the amplitude of connections and the directionality (Figure 3). The condition effects were also variable across methods (Table V). None of the effects were present

in all models and, in general, the significant differences across conditions varied in the path coefficient involved. As this is the preferred method of applying SEM to functional neuroimaging data [McIntosh, 1999], these results suggest caution with the condition effects interpretation, especially when confounding effects are present in the investigated data set.

No model could be found to fit all covariance matrices. It is not clear from these results whether this lack of consistency is due to intrinsic characteristics of fMRI data or to SEM sensitivity to sample specific variance when the modification indices are employed to reach a well-fitting model [Raykov and Widaman, 1995]. When one particular model was tested across the covariance matrices, for each covariance matrix the investigation of modification indices pointed to changes in the model that, when followed, produced an identical model as the one originally fitted for that particular sample. This observation suggests that the divergence in our models was not generated by the stepwise strategy to build the models but rather by the use of modification indices. It is expected that the resulting models will be sample specific whenever one relies on modification indices to reach a good fit [e.g., Gonçalves et al., 2001]. Three major reasons can explain the diversity of models observed in our study. First, the variation in preprocessing strategy and time course selection method may have introduced significant and irreconcilable changes in the functional connectivity structure. Second, the use of modification indices may have evolved the SEM estimation to a bifurcation point where a particular choice of model adjustment had a dramatic influence on the subsequent results. Finally, the lack of fit for one particular model across covariance matrices may have been induced by the overestimation of the χ^2 value due to deviations from multivariate normality. Although based on our results we cannot estimate the exact contribution of these three factors, it is important to emphasise that all models presented here are plausible and cannot be rejected on anatomical or cognitive grounds. This may be a typical result one can expect when investigating higher order cognitive brain functions characterised by weak anatomically constrained models. It is also important that the variability in connectivity estimates can be large with a robust and reliable task when the motion-correction strategy negligibly affected the statistical activation maps. Therefore, our results strongly suggest the importance of a very careful consideration of the sources of variance that can be associated with functional and effective connectivity in the fMRI data, as well as the benefits of a strongly constrained model.

Significant deviations from multivariate normality were observed both for the average and the single voxel time courses (Table IV, models D–G). These deviations are likely to be present in fMRI data regardless of the task. The SVD time courses were, surprisingly, multivariate normally distributed with all three motion-correction methods although the univariate distributions were significantly non-normal (since these time courses were estimated across conditions, a bimodal distribution was expected). Under these circum-

stances, the multivariate normality estimate could have been biased by the small sample size (90 observations). This result suggests the importance of both univariate and multivariate normality estimates. Due to deviations from multivariate normality, some path coefficients were probably overestimated. By inspection of the critical ratio (CR), path coefficients that were not significantly different from zero taking into account the deviation from multivariate normality (marked by a star in Figure 3) could be identified in some models. As Petersson et al. [1999] cautioned, an evaluation of the effects of assumption violation is crucial for the valid use of any statistical method. These violations do not imply that the method is incorrect, but rather establish some boundaries where the results can be trusted. For SEM applications to functional neuroimaging data, the assessment of multivariate normality is important in assessing the true significance of the path coefficients.

The estimated residual variances were also affected by the time course selection method and by the motion correction strategy (Table III). Fixing these parameters to some arbitrary values [McIntosh and Gonzales-Lima, 1994] can introduce sample dependencies for parameter estimates and the goodness of fit indices. The estimation of the residual variances within the model, when possible, is also contraindicated since it is better, in general, to fit models with a smaller number of parameters. As Hu and Bentler [1995] suggested, for complicated models with a large number of free parameters, a good fit can result from overparametrisation rather than from a correct model. Therefore, we recommend the objective estimation of the residual variances via the method proposed by Bullmore et al. [2000].

We do not believe that the variations in connectivity estimates presented here are specific to the OLR task. This study is intended to alert the neuroimaging community about the necessity of a more careful investigation of the data sets employed in connectivity studies. Until these problems are dealt with, it is very likely that any un-modelled source of noise or confounding effects could be dramatically reflected at the modelling stage. For higher order cognitive brain functions, due to the weak anatomical constraints, multiple models of effective connectivity can be integrated with the psychological theories and the anatomical information for the task under study. In conclusion, in order to avoid bias in the interpretation of the results, SEM should be used for fMRI data only with very strong constraints and even in these situations, the stability of the results with respect to preprocessing strategy and multivariate normality, in conjunction with the sample size, should be investigated. An important benefit can be achieved from methods such as transcranial magnetic stimulation [Fox et al., 1997] and electroencephalography that can be used in order to impose temporal constraints on the effective connectivity models.

ACKNOWLEDGMENTS

We express our gratitude to the three anonymous reviewers whose comments significantly improved the manuscript. We also thank Michael McStephen for very useful discus-

sions. We acknowledge support from the Neuroinformatics Platform funded by Neurosciences Victoria. M.G. acknowledges support of a scholarship from Monash University, Melbourne. G.S. acknowledges support of a fellowship from NARSAD, New York.

REFERENCES

- Beckmann C, Smith SM. 2002. Probabilistic Extensions to Independent Component Analysis in FMRI. Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan. Available on CD-Rom in NeuroImage, Vol. 16, No. 2.
- Büchel C, Friston KJ. 1997. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex* 7:768–778.
- Büchel C, Coull JT, Friston KJ. 1999. The predictive value of changes in effective connectivity for human learning. *Science* 283:1538–1541.
- Bullmore E, Brammer MJ, Rabe-Hesketh S, Curtis VA, Morris RG, Williams SCR, Sharma T, McGuire PK. 1999. Methods for diagnostic and treatment of stimulus-correlated motion in generic brain activation studies using fMRI. *Hum Brain Mapp* 7:38–48.
- Bullmore E, Horwitz B, Honey G, Brammer M, Williams S, Sharma T. 2000. How good is good enough in path analysis of fMRI data? *NeuroImage* 11:289–301.
- Bullmore E, Long C, Suckling J, Fadili J, Calvert G, Zelaya F, Carpenter TA, Brammer M. 2001. Colored noise and computational inference in neurophysiological (fMRI) time series analysis: resampling methods in time and wavelet domains. *Hum Brain Mapp* 12:61–78.
- Chou C-P, Bentler PM. 1995. Estimates and tests in structural equation modeling. In: Hoyle RH, editor. *Structural equation modeling: concepts, issues and applications*. Thousand Oaks, CA: Sage Publications. p 37–55.
- Christoff K, Whitfield SL, Shelton AL, Panitz DA, Gabrieli JDE. 2002. Comparison of BOLD hemodynamic responses across different brain regions. Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan. Available on CD-Rom in NeuroImage, Vol. 16, No. 2.
- Della-Maggiore V, Sekuler AB, Grady CL, Bennet PJ, Sekuler R, McIntosh AR. 2000. Corticolimbic interactions associated with performance on a short-term memory task are modified by age. *J Neurosci* 20:8410–8416.
- Espósito F, Formisano E, Seifritz E, Goebel R, Morrone R, Tedeschi G, Di Salle F. 2002. Spatial independent component analysis of functional MRI time-series: To what extent do results depend on the algorithm used? *Hum Brain Mapp* 16:146–157.
- Fletcher P, Büchel C, Josephs O, Friston KJ, Dolan R. 1999. Learning-related neural responses in prefrontal cortex studied with functional neuroimaging. *Cereb Cortex* 9:168–178.
- Fox P, Ingham R, George MS, Mayberg H, Ingham J, Roby J, Martin C, Jerabek P. 1997. Imaging human intra-cerebral connectivity by PET during TMS. *NeuroReport* 8:2787–2791.
- Freire L, Mangin J-F. 2001. Motion correction algorithms may create spurious brain activations in the absence of subject motion. *NeuroImage* 14:709–722.
- Friston KJ. 1994. Functional and effective connectivity in neuroimaging: A synthesis. *Hum Brain Mapp* 2:56–78.

- Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ. 1993a. Functional connectivity: The principal component analysis of large (PET) data sets. *J Cereb Blood Flow Metab* 13:5–14.
- Friston KJ, Frith CD, Frackowiak RSJ. 1993b. Time-dependant changes in effective connectivity measured with PET. *Hum Brain Mapp* 1:69–80.
- Friston KJ, Büchel C, Fink GR, Morris J, Rolls E, Dolan RJ. 1997. Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage* 6:218–229.
- Gavrilescu M, Shaw ME, Stuart GW, Eckersley P, Svalbe ID, Egan GF. 2002. Simulation of the effects of global normalization procedures in functional MRI. *NeuroImage* 17:532–542.
- Genovese R, Lazar NA, Nichols T. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* 15:870–878.
- Gonçalves MS, Hall DA, Johnsrude IS, Haggard MP. 2001. Can meaningful effective connectivities be obtained between auditory cortical regions? *NeuroImage* 14:1353–1360.
- Hampson M, Peterson BS, Skudlarski P, Gatenby JC, Gore JC. 2002. Detection of functional connectivity using temporal correlations in MR images. *Hum Brain Mapp* 15:247–262.
- Honey DG, Fu CHY, Kim J, Brammer MJ, Croudace TJ, Suckling J, Pich EM, Williams SCR, Bullmore ET. 2002. Effects of verbal working memory load on corticocortical connectivity modeled by path analysis of functional magnetic resonance imaging data. *NeuroImage* 17:573–582.
- Horwitz B. 1994. Data Analysis paradigm for metabolic-flow data: Combining neural modeling and functional neuroimaging. *Hum Brain Mapp* 2:112–122.
- Hoyle RH. 1995. The structural equation modeling approach: Basic Concepts and fundamental issues. In: Hoyle RH, editor. *Structural equation modeling: concepts, issues and applications*. Thousand Oaks, CA: Sage Publications. p 1–15.
- Hu L, Bentler PM. 1995. Evaluating model fit In: Hoyle RH, editor. *Structural equation modeling: concepts, issues and applications*. Thousand Oaks, CA: Sage Publications. p 76–99.
- Jenkinson M, Bannister P, Brady JM, Smith S. 2002. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 17:825–841.
- Jennings JM, McIntosh AR, Kapur S. 1998. Mapping neural interactivity onto regional activity: an analysis of semantic processing and response mode interactions. *NeuroImage* 7:244–254.
- Kendall M. 1994 In: Stuart A, Keith, J, editors. *Kendall's advanced theory of statistics*, 6th ed. New York: John Wiley and Sons.
- Krüger G, Glover GH. 2001. Physiological noise in oxygenation-sensitive magnetic resonance imaging. *Magn Reson Med* 46:631–637.
- Lowe MJ, Mock BJ, Soreson JA. 1998. Functional connectivity in single and multiple echoplanar imaging using resting-state fluctuations. *NeuroImage* 2:119–132.
- Lund TE. 2001. fMRI-Mapping functional connectivity or correlating cardiac-induced noise? *Magn Reson Med* 46:628.
- MacCallum RC. 1995. Model specification: Procedures, strategies, and related issues. In: Hoyle RH, editor. *Structural equation modeling: concepts, issues and applications*. Thousand Oaks, CA: Sage Publications. p 16–36.
- MacCallum RC, Wegener DT, Uchino BN, Fabrigar LR. 1993. The problem of equivalent models in applications of covariance structure analysis. *Psychol Bull* 114:185–199.
- Maguire EA, Mummary CJ, Büchel C. 2000. Patterns of hippocampal cortical interaction dissociate temporal lobe memory subsystems. *Hippocampus* 10:475–482.
- Mardia KV. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57:519–530.
- McIntosh AR. 1999. Mapping cognition to the brain through neural interactions. *Memory* 7:523–548.
- McIntosh AR, Gonzales-Lima F. 1994. Structural equation modeling and its application to network analysis in functional brain imaging. *Hum Brain Mapp* 2:2–22.
- McIntosh AR, Rajah MN, Lobaugh NJ. 1999. Interactions of prefrontal cortex in relation to awareness in sensory learning. *Science* 284:1531–1533.
- McKeown MJ, Makeig S, Brown SS, Jung T-P, Kindermann SS, Bell AJ, Sejnowski TJ. 1998. Analysis of fMRI data by blind separation into spatial independent component analysis. *Hum Brain Mapp* 6:160–168.
- Nyberg L, McIntosh AR, Cabeza R, Nilsson LG, Houle S, Habib R, Tulving E. 1996. Network analysis of positron emission tomography regional cerebral blood flow data: Ensemble inhibition during episodic memory retrieval. *J Neurosci* 11:3753–3759.
- Peltier SJ, and Noll DC. 2002. T_2^* dependence of low frequency functional connectivity. *NeuroImage* 16:985–992.
- Petersson KM, Nichols TE, Poline JB, Holmes AP. 1999. Statistical limitations in functional neuroimaging I. Non-inferential methods and statistical models. *Phil Trans R Soc Lond B* 354:1239–1260.
- Price CJ, Friston KJ. 1999. Scanning patients with tasks they can perform. *Hum Brain Mapp* 8:102–108.
- Raykov T, Widaman KF. 1995. Issues in applied structural equation modeling research. *Struct Equat Model* 2:289–318.
- Rowe J, Stephan KE, Friston KJ, Frackowiak R, Lees A, Passingham R. 2002. Attention to action in Parkinson's disease. Impaired effective connectivity among frontal cortical regions. *Brain* 125: 276–289.
- West SG, Finch JF, Curran PJ. 1995. Structural equation models with nonnormal variable: Problems and remedies. In: Hoyle RH, editor. *Structural equation modeling: concepts, issues and applications*. Thousand Oaks, CA: Sage Publications. p 56–75.
- Worsley KJ, Friston KJ. 1995. Analysis of fMRI time-series revisited-again. *NeuroImage* 2:173–181.
- Wright S. 1934. The method of path coefficients. *Ann Math Stat* 5: 161–215.
- Xiong J, Parsons LM, Gao J, Fox PT. 1999. Interregional connectivity to primary motor cortex revealed using MRI resting state images. *Hum Brain Mapp* 8:151–156.