Frequency domain simultaneous source and source coherence estimation with an application to MEG.

Grasman, R.P.P.P.; Huizenga, H.M.; Waldorp, L.J.; Böcker, K.B.E.; Molenaar, P.C.M.

Published in:
IEEE Transactions on Biomedical Engineering

DOI:
10.1109/TBME.2003.820385

Frequency Domain Simultaneous Source and Source Coherence Estimation With an Application to MEG

Raoul P. P. P. Grasman*, *Associate Member, IEEE*, Hilde M. Huizenga, Lourens J. Waldorp, *Associate Member, IEEE*, Koen B. E. Böcker, and Peter C. M. Molenaar

***Abstract*—Interactions between cortical areas are crucial for cognitive functioning. Methods currently in use to assess such interactions are not well suited for this task because they lack timing precision, localization precision, or both. We present a method for simultaneous estimation of source location and orientation parameters and cross-spectral parameters to overcome these problems. Different estimators are evaluated for their performance. From a simulation study, we conclude that the estimators derived from the maximum-likelihood procedure have good statistical properties with moderate sample sizes, whereas those obtained from the generalized least squares procedure are biased and give poor-quality standard errors. The method is illustrated with visually evoked field data, with inconclusive results**

***Index Terms*—Coherence, covariance structures, cross-spectrum structures, EEG, frequency domain, functional connectivity, MEG, source modeling.**

I. Introduction

An important goal of cognitive neuroimaging is to establish how different cortical areas cooperate and interact with each other to implement cognitive functions. As an example, in the last decade compelling evidence has become available on the role of the visual and parietal cortical areas and their interactions during the selective processing of visual stimuli (e.g., [1] and [2]). Other phenomena consistently found are increased synchronization of electric activity while viewing stimuli with emotional content [3], [4], and while performing cognitive tasks [5]. Such synchronizations of cortical rhythms are hypothesized to solve the binding problem in the perception of objects [6], [7].

Current techniques for localizing active cortical areas are functional magnetic resonance imaging (fMRI), positron emission tomography (PET), high-resolution cortical imaging [8], and equivalent current dipole (ECD) modeling of the magnetoencephalogram (MEG) and electroencephalogram (EEG) (e.g., [9]). Although advances have been made to establish cortical interactions with PET and fMRI (e.g., [10]–[12]), only an MEG-/EEG-based analysis has enough timing precision to be able to discern mutual influences between activities of different cortical areas directly.

Current commonly applied methods to extract information on interactions between different brain regions from EEG/MEG rely on the analysis of raw sensor signals. Standard methods include coherence analysis [13]–[16] and event-related synchronization/(de-)synchronization (ERS/ERD) [17]. Coherence is assumed to be indicative of interactions between spatially separated areas, whereas ERS/ERD is interpreted as a measure of small scale interactions within brain areas (see [17]). High coherence between sensors is often attributed to coherent activity of sources directly beneath these sensors. This interpretation is not generally valid. Problems with this interpretation of MEG/EEG coherence lie in the fact that all sensors are sensitive to sources everywhere in the head. Moreover, for EEG there are additional confounding influences: volume conduction effects [18], [19], and the reference effect. Therefore, it is very difficult to localize areas responsible for the observed coherence [20]. ERS and ERD suffer from similar problems [21]. As an illustration, Fig. 1(a) depicts a traditional sensor coherence analysis obtained from simulated data. The coherence pattern suggests that sources in the temporal and parietal cortical regions in the right hemisphere were correlated, and that possibly left and right hemispheric temporal cortical regions were correlated. In contrast, Fig. 1(b) shows the dipoles that were used to generate the coherence pattern in Fig. 1(a); a dipole in the parietal lobe of each hemisphere. In general, it will, thus, be as difficult to localize coherent cortical areas by mere inspection of coherences between sensors, as it is to localize cortical areas by mere inspection of iso-field/potential contour plots.

ECD modeling allows sources to be localized from the MEG/EEG. Currently, only source locations, orientations, and trial mean amplitudes are estimated from trial-averaged data. This, however, disregards the information that is present in the amplitude variation across trials (but see [22]). Suppose for instance that there are two sources with partly stochastic amplitudes. Then there will be variation of the source amplitudes across trials, and possibly covariation if the sources are interacting. Information about this covariation of sources is, thus, contained in single-trial MEG/EEG signals. In the currently presented method, this information is exploited to

Manuscript received June 19, 2002; revised April 18, 2003. This work was supported by The Netherlands Organization for Scientific Research (NWO) through the NWO foundation for Behavioral and Educational Sciences under a grant awarded to H. M. Huizenga, and P. C. M. Molenaar. The work of H. M. Huizenga was also supported in part by a KNAW fellowship. The work of R. P. P. P. Grasman was supported in part by the NWO under Grant 527-25-014. The work of L. J. Waldorp was supported by the NWO under Grant 527-25-013. The work of K. B. E. Böcker was supported by the NWO under Grant 527-25-015.

*R. P. P. P. Grasman is with the Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018WB Amsterdam, The Netherlands (e-mail: grasman@psy.uva.nl).

H. M. Huizenga, L. J. Waldorp, and P. C. M. Molenaar are with the Department of Psychology, University of Amsterdam, 1018WB Amsterdam, The Netherlands.

K. B. E. Böcker is with the Department of Psychopharmacology, Utrecht University, 3583CA Utrecht, The Netherlands.

Digital Object Identifier 10.1109/TBME.2003.820385

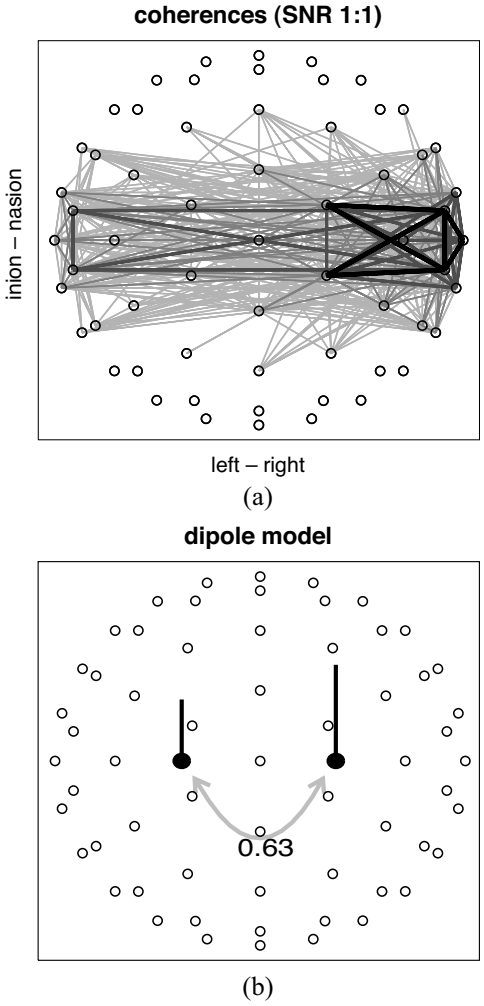

Fig. 1. (a) Coherence analysis of simulated data that were obtained as described in Section III. Coherences between sensors ranging between 0.1 and 0.52, are indicated by a line joining them. Thicker and blacker lines indicate higher coherence. (b) Aactual dipole locations, orientations, and amplitude coherence (indicated by the gray two-way arrow) and relative power (indicated by the length of the dipole moments—no units attached). The coherences were averaged across the included frequencies in $\Omega^{\#}$. See the text for a definition of $\Omega^{\#}$.

determine covariation between sources. This then overcomes the problems associated with the aforementioned techniques, by directly estimating source locations, source orientations, and source covariances. The estimation procedure can be carried out in the frequency domain (see, [23]–[25]) which, under the assumption of stationarity, can greatly reduce the computational burden as will be argued in Section II.

II. Method

A. Model Specification

Let $y_{m,l}(t)$ denote the signal measured at the mth sensor at time t in the lth trial sampled with fixed sampling frequency, and define the vector $\mathbf{y}_l(t) = [y_{1,l}(t), \ldots, y_{M,l}(t)]^T$ as the array of measurements made by the M sensors at times $t = 1, \ldots, N$ in trials $l = 1, \ldots, L$. Here, ^{T} denotes transposition. The matrix $\mathbf{Y}_l = [\mathbf{y}_l(1), \ldots, \mathbf{y}_l(N)]$ will then denote the $M \times N$ data matrix measured in trial l. We use the spatiotemporal dipole analysis model in which we assume a total of D fixed dipoles generating the data in $\mathbf{Y}_l$

$$\mathbf{Y}_l = \mathbf{G}(\theta)\Xi_l + \mathbf{E}_l. \tag{1}$$

Here, $\mathbf{G}(\theta)$ is the $M \times D$ gain matrix with unit activity of D dipoles, depending on the location parameters and orientation parameters θ which are fixed over time samples and trials, $\boldsymbol{\Xi}_l$ is the $D \times N$ matrix containing the source amplitude time series in trial l, and $\mathbf{E}_l$ is the $M \times N$ matrix of noise signals on trial l. The parameter θ consists of locations θ_d^{loc} and orientations θ_d^{or} for $d = 1, \ldots, D$. If we let $\mathrm{vec}\{\mathbf{Y}_l\}$ denote the operator that stacks the columns of $\mathbf{Y}_l$ in a column vector, then with the Kronecker product $\otimes$ and the equality $\mathrm{vec}\{ABC\} = (C^T \otimes A)\mathrm{vec}\{B\}$ [26, p. 30], the model can be written as

$$\mathbf{y}_l = [\mathbf{I}_N \otimes \mathbf{G}(\theta)]\,\boldsymbol{\eta}_l + \boldsymbol{\varepsilon}_l \tag{2}$$

where $\mathbf{y}_l = \mathrm{vec}\{\mathbf{Y}_l\} = [\mathbf{y}_l(1)^T, \ldots, \mathbf{y}_l(N)^T]^T$ is the $M \cdot N$ vector containing the measurements in trial l. Similarly, $\boldsymbol{\eta}_l = \mathrm{vec}\{\boldsymbol{\Xi}_l\}$ is the $D \cdot N$ vector that contains the source amplitudes, and $\boldsymbol{\varepsilon}_l = \mathrm{vec}\{\mathbf{E}_l\}$ is the $M \cdot N$ vector with noise signals. We assume that the noise signals $\boldsymbol{\varepsilon}_l$ have zero mean. Note that we use Greek symbols to denote unobserved quantities.

In this paper, we will focus only on source coherences, and we will disregard the mean amplitude. Since we are interested in covariation between sources and not in trial source amplitude time functions, we may consider only the covariance matrix implied by (2). Let $\boldsymbol{\Sigma}$ denote the covariance matrix of $\mathbf{y}_l$, $\boldsymbol{\Psi}$ the covariance matrix of $\boldsymbol{\eta}_l$, and $\boldsymbol{\Theta}$ the covariance matrix of $\boldsymbol{\varepsilon}_l$. Then, from (2) and the additional assumption that the source signals $\boldsymbol{\eta}_l$ and noise signals $\boldsymbol{\varepsilon}_l$ are uncorrelated we have

$$\begin{aligned}\boldsymbol{\Sigma} &= \langle (\mathbf{y}_l - \langle \mathbf{y}_l \rangle)(\mathbf{y}_l - \langle \mathbf{y}_l \rangle)^T \rangle \\ &= [\mathbf{I}_N \otimes \mathbf{G}(\theta)]\,\boldsymbol{\Psi}\,[\mathbf{I}_N \otimes \mathbf{G}(\theta)]^T + \boldsymbol{\Theta}\end{aligned} \tag{3}$$

where $\langle \cdot \rangle$ indicates the ensemble average. The matrix $\boldsymbol{\Sigma}$ is prohibitively large to estimate by conventional means, containing $M \cdot N(M \cdot N + 1)/2$ unique elements (much more than the number of trials L in most situations).

The situation can be significantly simplified, however, if we Fourier transform the data, making a further assumption that the variation of the source amplitudes around their ensemble averages $\langle \boldsymbol{\eta}_l \rangle$ and noise signals are stationary and satisfy certain mixing conditions [23] and [27]: Define $\widetilde{\mathbf{y}}_l(k) = (1/2\pi N)^{1/2} \sum_t \mathbf{y}_l(t) \exp(-i2\pi tk/N)$ to be the discrete Fourier transform coefficient at frequencies $2\pi k/N$, $k = 1, \ldots, K < N/2$ [23], and let $\widetilde{\mathbf{y}}_l = [\widetilde{\mathbf{y}}_l^T(1), \ldots, \widetilde{\mathbf{y}}_l^T(K)]^T$. Define $\widetilde{\boldsymbol{\eta}}_l(k)$ and $\widetilde{\boldsymbol{\varepsilon}}_l(k)$ and $\widetilde{\boldsymbol{\eta}}_l$ and $\widetilde{\boldsymbol{\varepsilon}}_l$ similarly. Throughout the rest of this paper $\widetilde{\ }$ will indicate frequency domain quantities. One advantage of working in the frequency domain (cf. [23]) arises in that different frequencies have asymptotically uncorrelated Fourier coefficients, i.e., whereas $\langle \boldsymbol{\varepsilon}_l(t)\boldsymbol{\varepsilon}_l^T(s) \rangle$ does not necessarily equal zero for $t \neq s$, $\langle \widetilde{\boldsymbol{\varepsilon}}_l(j)\widetilde{\boldsymbol{\varepsilon}}_l^*(k) \rangle \to 0$ for $j \neq k$ as $N \to \infty$ [27], where * denotes complex conjugation and transposition. Similarly, $\langle [\widetilde{\mathbf{y}}_l(k) - \langle \widetilde{\mathbf{y}}_l(k) \rangle][\widetilde{\mathbf{y}}_l(j) - \langle \widetilde{\mathbf{y}}_l(j) \rangle]^* \rangle \to 0$

and $\langle[\widetilde{\boldsymbol{\eta}}_l(k) - \langle\widetilde{\boldsymbol{\eta}}_l(k)\rangle][\widetilde{\boldsymbol{\eta}}_l(j) - \langle\widetilde{\boldsymbol{\eta}}_l(j)\rangle]^*\rangle \to 0$ when $k \neq j$ as $N \to \infty$. This implies that the covariance matrices $\widetilde{\boldsymbol{\Sigma}}$, $\widetilde{\boldsymbol{\Psi}}$, and $\widetilde{\boldsymbol{\Theta}}$ of $\widetilde{\mathbf{y}}_l$, $\widetilde{\boldsymbol{\eta}}_l$, and $\widetilde{\boldsymbol{\varepsilon}}_l$, respectively, are approximately block diagonal for large N. Denoting the model in (2) in the frequency domain $\widetilde{\mathbf{y}}_l = [\mathbf{I}_K \otimes \mathbf{G}(\theta)]\widetilde{\boldsymbol{\eta}}_l + \widetilde{\boldsymbol{\varepsilon}}_l$ (see, also, [23] and [25]), we find its covariance matrix to be

$$\begin{aligned}\widetilde{\boldsymbol{\Sigma}} &= [\mathbf{I}_K \otimes \mathbf{G}(\theta)]\,\widetilde{\boldsymbol{\Psi}}\,[\mathbf{I}_K \otimes \mathbf{G}(\theta)]^T + \widetilde{\boldsymbol{\Theta}} \\ &\approx \begin{bmatrix} \widetilde{\boldsymbol{\Sigma}}_1 & 0 & \cdots & 0 \\ 0 & \widetilde{\boldsymbol{\Sigma}}_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \widetilde{\boldsymbol{\Sigma}}_K \end{bmatrix}\end{aligned} \tag{4}$$

where $\widetilde{\boldsymbol{\Sigma}}_k = \mathbf{G}(\theta)\widetilde{\boldsymbol{\Psi}}_k\mathbf{G}(\theta)^T + \widetilde{\boldsymbol{\Theta}}_k$, $k = 1, \ldots, K$. The $D \times D$ matrices $\widetilde{\boldsymbol{\Psi}}_k$ and the $M \times M$ matrices $\widetilde{\boldsymbol{\Theta}}_k$ are the block-diagonal elements of $\widetilde{\boldsymbol{\Psi}}$ and $\widetilde{\boldsymbol{\Theta}}$, respectively. The elements of $\widetilde{\boldsymbol{\Sigma}}_k$ are the power- and cross-spectra of the MEG/EEG signals [23]. The elements of $\widetilde{\boldsymbol{\Psi}}_k$ are the power- and cross-spectra of the source-amplitude signals and, therefore, provide a measure of the strength of interactions between the sources; $\widetilde{\boldsymbol{\Theta}}_k$ similarly specify the power- and cross-spectra of the noise signals. Compared with conventional coherence analysis, an estimate of $\widetilde{\boldsymbol{\Psi}}_k$ has the advantage of describing covariation of cortical areas in terms of amplitude cross-spectra of sources.

B. Parameter Estimation

Let $\widetilde{\mathbf{V}}$ denote the matrix

$$\widetilde{\mathbf{V}} = \begin{bmatrix} \widetilde{\mathbf{V}}_1 & 0 & \cdots & 0 \\ 0 & \widetilde{\mathbf{V}}_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \widetilde{\mathbf{V}}_K \end{bmatrix} \tag{5}$$

where $\widetilde{\mathbf{V}}_k = 1/(L-1)\sum_{l=1}^{L}[\widetilde{\mathbf{y}}_l(k) - \widetilde{\mathbf{y}}(k)][\widetilde{\mathbf{y}}_l(k) - \widetilde{\mathbf{y}}(k)]^*$ and $\widetilde{\mathbf{y}}(k) = (1/L)\sum_l \widetilde{\mathbf{y}}_l(k)$ the mean across trials.[1] $\widetilde{\mathbf{V}}_k$ is the observed cross-spectrum, and is an approximately unbiased estimate of the cross-spectrum $\widetilde{\boldsymbol{\Sigma}}_k$ of (4) as $N \to \infty$ for $k = 1, \ldots, K$ [27, ch. 7].

Collecting the unknown parameters of $\widetilde{\boldsymbol{\Psi}}$ and $\widetilde{\boldsymbol{\Theta}}$, together with the unknown parameters in θ, in the p-dimensional vector ξ, then from (4) we obtain

$$\widetilde{\boldsymbol{\Sigma}}(\xi) = [\mathbf{I}_K \otimes \mathbf{G}(\xi)]\,\widetilde{\boldsymbol{\Psi}}(\xi)\,[\mathbf{I}_K \otimes \mathbf{G}(\xi)]^T + \widetilde{\boldsymbol{\Theta}}(\xi). \tag{6}$$

The true value of ξ will be denoted by ξ_0, i.e., $\widetilde{\boldsymbol{\Sigma}}(\xi_0) \equiv \widetilde{\boldsymbol{\Sigma}}$ of (4).

To estimate the model parameters ξ, a discrepancy measure between $\widetilde{\boldsymbol{\Sigma}}(\xi)$ and $\widetilde{\mathbf{V}}$ can be minimized. Standard discrepancy measures have the least squares form [28], [29]

$$F(\xi) = \frac{1}{2}\left\|\mathbf{A}\left[\widetilde{\mathbf{V}} - \widetilde{\boldsymbol{\Sigma}}(\xi)\right]\mathbf{A}^*\right\|_F^2 \tag{7}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. $\mathbf{A}$ can be any weight matrix such that $\mathbf{A}^*\mathbf{A}$ is positive definite, but we will only consider cases in which it has a similar blockdiagonal structure as $\widetilde{\mathbf{V}}$. Particular choices of $\mathbf{A}$ are $\mathbf{I}_{K\cdot M}$ and $\widetilde{\mathbf{V}}^{-1/2}$, where $\mathbf{I}_{KM}$ is the $K \cdot M$ identity matrix and $^{-1/2}$ denotes the inverse of the Cholesky factor of $\widetilde{\mathbf{V}}$ [30], resulting in unweighted (ULS) and generalized (GLS) least squares parameter estimates, respectively [28]. The loss function with these particular choices of $\mathbf{A}$ will be denoted $F_{\mathrm{ULS}}(\xi)$ and $F_{\mathrm{GLS}}(\xi)$, respectively, throughout this paper. In [31] and [32], it was shown that GLS estimators asymptotically have the lowest variance compared with all other choices of $\mathbf{A}$. Moreover they are efficient if $\widetilde{\mathbf{V}}$ follows a Wishart distribution law (in the sense that it is equal to the Cramér-Rao lower bound for the error variance of any unbiased estimator [33]). In fact, due to a central limit theorem for the Fourier coefficients, $\widetilde{\mathbf{V}}_k$ is asymptotically distributed as a complex Wishart variable for all $k = 1, \ldots, K$, even when the signals themselves are not Gaussian [27]. This is another advantage of working in the frequency domain (see, also, [23]).

The approximate complex Wishart distribution of $\widetilde{\mathbf{V}}_k$ also allows the parameters to be estimated using the (approximate) maximum-likelihood (ML) method: The ML-estimation function can then be defined as (cf. [31] and [34])

$$F_{\mathrm{ML}}(\xi) = -p + \ln|\widetilde{\boldsymbol{\Sigma}}(\xi)| + \mathrm{tr}\{\widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Sigma}}(\xi)^{-1}\} - \ln|\widetilde{\mathbf{V}}|. \tag{8}$$

Here, $\mathrm{tr}\{\cdot\}$ gives the trace of a matrix, and $|\cdot|$ the determinant. The constants p and $\ln|\widetilde{\mathbf{V}}|$ are of course not essential for the optimization of (8). These ML estimators are also asymptotically efficient [31].

Taking advantage of the block–diagonal structure of $\widetilde{\mathbf{V}}$, $\widetilde{\boldsymbol{\Sigma}}(\xi)$ and $\mathbf{A}$, both (7) and (8) can be rewritten as a sum over frequencies. As an additional advantage, merely a subset of frequencies can be used, effectively filtering out frequencies not of interest, and increasing the signal-to-noise ratio (SNR) if the source amplitudes have zero power at these omitted frequencies. Denoting this subset with $K^{\#}$ included frequencies by $\Omega^{\#}$, we have

$$F(\xi) = \frac{1}{2}\sum_{k\in\Omega^{\#}} \|\mathbf{A}_k[\widetilde{\mathbf{V}}_k - \widetilde{\boldsymbol{\Sigma}}_k(\xi)]\mathbf{A}_k^*\|_F^2$$

and

$$F_{\mathrm{ML}}(\xi) = -p + \sum_{k\in\Omega^{\#}}[\ln|\widetilde{\boldsymbol{\Sigma}}_k(\xi)| + \mathrm{tr}\{\widetilde{\mathbf{V}}_k\widetilde{\boldsymbol{\Sigma}}_k(\xi)^{-1}\} - \ln|\widetilde{\mathbf{V}}_k|]$$

where $\mathbf{A}_k$ are the $M \times M$ block-diagonal elements of $\mathbf{A}$. Minimization of either $F_{\mathrm{ML}}(\xi)$ or $F(\xi)$ with some choice of $\mathbf{A}$ yields the desired estimates of θ, $\widetilde{\boldsymbol{\Psi}}_k$ and $\widetilde{\boldsymbol{\Theta}}_k$, for $k \in \Omega^{\#}$.

Let each dipole be specified by r parameters. The number of unique parameters of ξ is $r \cdot D$ dipole parameters, plus $D^2K^{\#}$ for the nonzero real and imaginary parts of $\widetilde{\boldsymbol{\Psi}}_k$ plus $M^2K^{\#}$ parameters for the nonzero real and imaginary parts of $\widetilde{\boldsymbol{\Theta}}_k$, $k \in \Omega^{\#}$. There are, therefore, many more parameters than the $M^2K^{\#}$ available nonduplicate values in $\widetilde{\mathbf{V}}_k$, $k \in \Omega^{\#}$. Hence, without further assumptions, i.e., without a reduction of the number of free parameters in the model, the parameter vector ξ is not identifiable. Therefore, assumptions on $\widetilde{\boldsymbol{\Psi}}$ or $\widetilde{\boldsymbol{\Theta}}$ (or both) have to be made. The assumption that we will adopt here is that $\widetilde{\boldsymbol{\Theta}}_k = \widetilde{\sigma}_k^2\mathbf{I}_M$, that is, the noise signals are uncorrelated between

[1]For $\boldsymbol{\eta}_l$ and $\boldsymbol{\varepsilon}_l$ stationary, $\widetilde{\mathbf{y}}(k)$ should be close to zero for all $k = 1, \ldots, K$ [27] and is immaterial.

sensors. If this is not a valid assumption, and the sensor noise covariance $\mathbf{U}$ is (approximately) known up to a multiplicative constant (e.g., from measurements prior to the experiment), then $\widetilde{\mathbf{\Theta}}_k = \widetilde{\sigma}_k^2 \mathbf{U}$ can be taken.

The parameters ξ may be required to satisfy q constraints

$$\mathbf{r}(\xi) = \mathbf{0} \tag{9}$$

where $\mathbf{r}(\xi)$ is a real continuously differentiable q-vector valued function [35]. This occurs for example in symmetric dipole pair models (e.g., [36] and [37]). It also occurs where it is required that sources are tangential, for example in a MEG analysis in a spherical head model. More specifically, if both the location θ_d^{loc} and orientation θ_d^{or} of a dipole is specified in Euclidean coordinates and θ_d^{loc} has its origin at the center of the sphere, and θ_d^{or} has its origin at θ_d^{loc}. Then, the constraint that the dipole orientation vector θ_d^{or} is orthogonal to the dipole location vector θ_d^{loc} has to be imposed: $(\theta_d^{\mathrm{loc}})^T \theta_d^{\mathrm{or}} \equiv 0$ [35], [38]. Then, $F(\xi)$ or $F_{\mathrm{ML}}(\xi)$ has to be minimized subject to (9).

The GLS and ML estimates have asymptotically a multivariate normal distribution [31]. An approximation to the covariance matrix of $\hat{\xi}$ is $\hat{\mathbf{B}} = \mathbf{H}(\hat{\xi})^{-1}/L$, where

$$\mathbf{H}(\hat{\xi}) = \left.\frac{\partial^2 F_{\mathrm{GLS}}(\xi)}{\partial \xi \partial \xi^T}\right|_{\xi=\hat{\xi}} \quad \text{or} \quad \mathbf{H}(\hat{\xi}) = \left.\frac{\partial^2 F_{\mathrm{ML}}(\xi)}{\partial \xi \partial \xi^T}\right|_{\xi=\hat{\xi}}. \tag{10}$$

The quality of this approximation, however, highly depends on the number L of trials and on the nonlinearity of the loss function, and has to be assessed by simulation. If there are any constraints on the parameters ξ, it was derived in [33] and [39] that the approximate covariance matrix of $\hat{\xi}$ may be obtained from the upper left $p \times p$ submatrix $\hat{\mathbf{B}}$ of

$$\frac{1}{L}\begin{bmatrix} \mathbf{H}(\hat{\xi}) + \mathbf{R}(\hat{\xi})^T \mathbf{R}(\hat{\xi}) & \mathbf{R}(\hat{\xi})^T \\ \mathbf{R}(\hat{\xi}) & \mathbf{0} \end{bmatrix}^{-1} \tag{11}$$

where $\mathbf{R}(\xi) = ((\partial \mathbf{r}(\xi))/(\partial \xi^T))$ is the $q \times p$ Jacobian matrix of the constraints $\mathbf{r}(\xi)$; see, [40] for a full discussion on the properties of the matrix in (11). The diagonal elements from the matrix $\hat{\mathbf{B}} = (\hat{b}_{ik})$ in (11) can be used to construct confidence intervals for the parameter estimates, by using the fact that the estimates $\hat{\xi}_i$, $i = 1, \ldots, p$ have an asymptotic normal distribution with standard deviation $\sqrt{\hat{b}_{ii}}$ [31], [40]. Then, the $(1-\alpha) \times 100\%$ confidence interval is given by $(\hat{\xi}_i - \Phi^{-1}(1-\alpha/2)\sqrt{\hat{b}_{ii}}, \hat{\xi}_i + \Phi^{-1}(1-\alpha/2)\sqrt{\hat{b}_{ii}})$, where Φ^{-1} denotes the inverse of the normal distribution function. A conventional choice of α is $\alpha = 0.05$.

Misspecifications in the model in (6) may result in considerably biased estimators. Apart from errors in the head model, incorrect assumptions on the noise structure and incorrect number of sources D are major sources of misspecification. To evaluate the fitted model we will use the chi-square goodness of fit test approach. In [31], it was shown that if the model is correct and L is large, then $L \cdot F_{\mathrm{GLS}}(\hat{\xi})$ is approximately a χ^2 variable with $M^2 K^{\#} - p + q$ degrees of freedom. $L \cdot F_{\mathrm{ML}}(\hat{\xi})$ has the same approximate χ^2 distribution [41]. Because the ML statistic is known to perform poorly with small numbers of trials, the Bartlett corrected statistic $\gamma(L-1)F_{\mathrm{ML}}(\hat{\xi})$ is used, where $\gamma = \{1 - [2M + 1 - 2/(M+1)]/6(L-1)\}$ is a factor that reduces deviation from the asymptotic chi-square behavior of this statistic [41]. Both statistics can be used to test the null hypothesis that the model is correct and can, therefore, be helpful in determining how many dipole sources are necessary to give an accurate description of the data. Other model selection procedures were discussed in [42].

III. Simulations

A first set of simulations was carried out to assess and compare the performance of both the GLS and the ML method when used with a realistic number of trials L and signal length N. The performance of the parameter estimators was assessed by testing whether they are unbiased, have good approximate standard errors $\sqrt{\hat{b}_{ii}}$, and are approximately normally distributed. The combined effect of violations of these asymptotic properties is condensed in an evaluation of the coverage rates of the confidence intervals. The effect of different levels of noise was assessed by comparing two noise conditions. A second, separate set of simulations was conducted to assess to what extend the chi-square goodness of fit was helpful in determining the number of sources. The purpose was to see if the chi-square statistic would reject an (incorrect) model with two dipoles in favor of a (correct) model with three dipoles, when in fact three dipoles were used to generate the data.

A. Data Generation

MEG data were generated using a unit spherical head model. In the first set of simulations, two current dipoles were situated at the points (0.0, 0.5, 0.75) and (0.0, −0.5, 0.75) (origin at the center of the sphere), both having the orientation (1,0,0), where the origin of the orientation frame of reference is situated at the dipole's location. Coherence of the source amplitudes was induced by letting them follow a first-order vector autoregressive stochastic process (VAR(1) process) [27]: $\eta_1(t) = 0.7\,\eta_1(t-1) + a_1(t)$ and $\eta_2(t) = 0.5\,\eta_1(t-1) + 0.7\,\eta_2(t-1) + a_2(t)$ where $a_1(t), a_2(t) \sim N(0,1)$, independent across time and trials. For the simulations in which the performance of the chi-square goodness of fit was determined, an additional source was located at (−0.5, 0.0, 0.75) with orientation (0,1,0), and amplitude $\eta_3(t) = 0.3\,\eta_1(t-1) + 0.7\,\eta_3(t-1) + a_3(t)$, again with $a_3(t) \sim N(0,1)$.

MEG was calculated for first-order gradiometers at 61 different locations evenly spread across the head (one gradiometer right above the vertex and four rings around it, spaced by 22.5°, containing 6, 12, 18, and 24 gradiometers, respectively). Autocorrelated noise [AR(1) process] was added to each gradiometer measurement. These noise processes were uncorrelated between MEG sensors. The SNR, defined as the ratio between the largest mean standard deviation of the (noiseless) sensor signals and mean standard deviation of the noise signals, was chosen to be[2]

1:5 and 1:1. In total, 400 trials were generated this way, each trial consisting of 128 samples, sampled at 128 Hz. The data were Fourier transformed using a fast Fourier transform algorithm, and the sample cross-spectral matrices $\widetilde{\mathbf{V}}_k, k \in \Omega^{\#}$, were calculated as indicated after (5).

B. Parameter Estimation

In the estimation problem, the dipole orientation parameters θ^{or} were specified in Cartesian coordinates. This requires the nonlinear constraint that the orientation vectors have unit norm, i.e., $\|\theta_d^{\text{or}}\| \equiv 1$ for $d = 1, \ldots, D$, where d indexes the dipole. In addition, we have the aforementioned orthogonality constraints on the dipole location and orientation parameters $(\theta_d^{\text{loc}})^T \theta_d^{\text{or}} \equiv 0, d = 1, \ldots, D$. In the first set of simulation we, therefore, have $2 \times 6 = 12$ dipole parameters subject to four constraints. In the second set of simulations, we also have $3 \times 6 = 18$ subject to six constraints for the three dipoles fit.

For the first set of simulations, a subset of ten frequencies, which contained most of the power in the generated signals, was selected for estimation: $\Omega^{\#} = \{4, \ldots, 13\}$ (in Hertz.). For two dipoles, $\widetilde{\mathbf{\Psi}}_k$ has four, and $\widetilde{\mathbf{\Theta}}_k = \widetilde{\sigma}_k^2 \mathbf{I}_M$ has one unknown parameters for each $k \in \Omega^{\#}$. This totals $10 \times (4 + 1) = 50$ parameters. We, therefore, have $p = 12 + 50 = 62$ parameters in the first set of simulations. In the second set of simulations, the fit was limited to a subset of 5 frequencies $\Omega^{\#} = \{4, \ldots, 8\}$. We, therefore, have $p = 12 + 5 \times (4 + 1) = 37$ parameters for two sources, and $p = 18 + 5 \times (9 + 1) = 68$ parameters for three sources in this simulation set.

Estimates $\hat{\xi}$ were obtained by minimizing $F_{\text{GLS}}(\xi)$ or $F_{\text{ML}}(\xi)$ under the indicated constraints, using a quasi-Newton algorithm [43]. Approximate standard errors were obtained from (11), where $\mathbf{H}(\hat{\xi})$ was approximated by a finite difference from the gradient of the loss function: Let e_j be the vector with all entries equal to zero, except the jth entry which is equal to 1, let $\delta_i = \epsilon \max(1, |\hat{\xi}_i|)$ where $\epsilon > 0$ is some small number (e.g., $\epsilon = 10^{-8}$). An approximate Hessian matrix of $F_{\text{GLS}}(\xi)$ was obtained from $\mathbf{H}(\hat{\xi}) \approx ((\nabla_i F_{\text{GLS}}(\hat{\xi} + \delta_j e_j) - \nabla_i F_{\text{GLS}}(\hat{\xi}))/\delta_j)$, where $\nabla_i = \partial/\partial\xi_i$, [40]. A similar approximation was obtained for $F_{\text{ML}}(\xi)$. It was found in [44] and [35] that such approximations are adequate.

In the second set of simulations, first a two-dipole model was fitted and, subsequently, a three-dipole model. The chi-square test was then used to see if the two-dipole model was rejected, and the three-dipole model was accepted. The SNR in this set of simulations was 1:1. The simulations were repeated 300 times for each combination of SNR (1:1 and 1:5) and estimation method (GLS and ML). The second set of simulations was only carried out for ML and an SNR of 1:1. The simulations in this set were also repeated 300 times.

C. Simulation Results

Bias was assessed by testing the null-hypothesis $H_0 : \hat{\xi}_\cdot = \xi_0$, where $\hat{\xi}_\cdot$ denotes the average of $\hat{\xi}$ across the 300 simulations, against the alternative $H_A : \hat{\xi}_\cdot \neq \xi_0$ using a Hotelling's T^2 test. Post hoc t-tests were used to locate the source of bias to specific (types of) parameters. The quality of the estimated standard errors was assessed with a chi-square test, comparing the variances of the ratios $(\hat{\xi}_i - (\hat{\xi}_\cdot)_i)/\sqrt{\hat{b}_{ii}}$ with their asymptotic expected value 1. For each estimator in $\hat{\xi}$ a Kolmogorov–Smirnov test was used to assess whether they had a normal distribution. As the orientation parameters satisfy constraints, their distribution is improper (i.e., their covariance matrix is singular). It should, therefore, be expected that the standard errors obtained for these parameters will be less than adequate.

The overall bias test indicated a clear difference between $\hat{\xi}_\cdot$ and ξ_0 in all simulations (ML 1:1, ML 1:5, GLS 1:1, and GLS 1:5). The source of this bias could be located to the different types of parameters.

In all conditions, the location and orientation estimators were essentially unbiased.

The $\widetilde{\mathbf{\Theta}}$ and $\widetilde{\mathbf{\Psi}}$ parameters, on the other hand, were statistically significantly biased. In Fig. 2, the mean parameters are plotted together with their true values. For the GLS estimates the bias was rather large; the $\widetilde{\mathbf{\Psi}}$ and $\widetilde{\mathbf{\Theta}}$ parameters were underestimated on average well over 25% of their true value. For ML, $\widetilde{\mathbf{\Psi}}$ and $\widetilde{\mathbf{\Theta}}$ bias was small: on average within approximately 2% of their true values. Different SNRs did not seem to affect bias significantly.

In Table I, the ratio between the mean of the estimated standard errors of the 300 simulations (estimated standard errors) and the standard deviations of the parameter estimates determined from the 300 simulations (simulation standard errors) is given, averaged over parameters. Values close to one indicate good agreement between the estimated and simulation standard errors. As can be seen in Table I, the ML estimated standard errors are good. This was confirmed by the chi-square tests: none of the ML estimated standard errors deviated significantly from their asymptotic value, whereas all of the GLS estimated standard errors did. It is known that dipole estimation algorithms generally have more difficulty in determining the radial component of dipole locations than the tangential components [38]. This was verified by testing whether the variance of the location estimates along the radial direction was larger than along tangential directions. This was indeed the case for all types of estimates; the variance of the location estimate was more than two times larger in the radial direction than in the tangential directions. For completeness, the root mean squared errors (RMSEs) for all types of parameters are given in Table II. It is seen that in all cases the RMSEs of GLS estimators are larger than those of the corresponding ML estimators. This is partly due to larger bias of the former, and partly due to larger variances of these estimators.

In all cases, the θ^{loc} estimators were normally distributed, as indicated by the Kolomogorov–Smirnov tests. The θ^{or} estimators were not, due to the constraints. Both GLS and ML $\widetilde{\mathbf{\Psi}}$ estimators were not normally distributed as indicated by the tests, but the deviance from normality was very small as can be seen in Fig. 3. The $\widetilde{\mathbf{\Theta}}$ estimators were normally distributed in all cases.

These results may justify the use of parameter confidence regions given in Section II-B in the case of finite sample ML es-

[2]The SNR was chosen to be consistent with ERPs and background EEG bandwidths found in the literature: ERP typically tend to have minimum to maximum amplitude bandwidths of about 5–15 μV, whereas background EEG tends to vary in bandwidths of 50–100 μV.

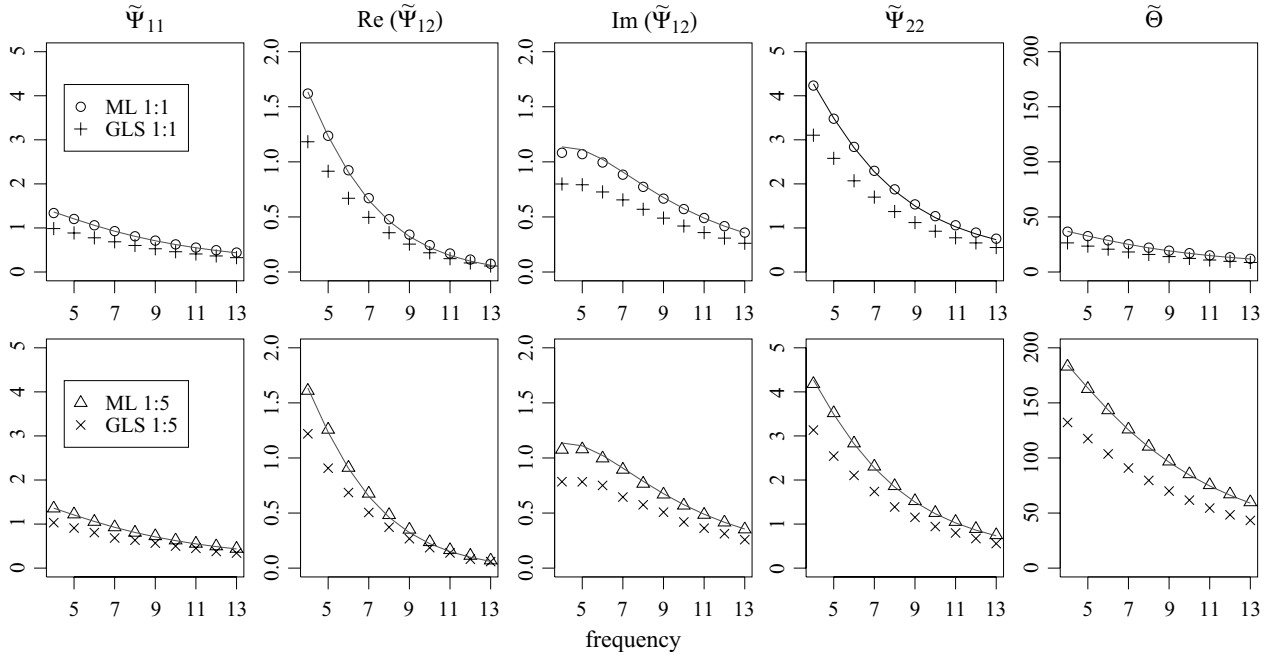

Fig. 2. Parameter bias. Simulation means of the $\widetilde{\boldsymbol{\Psi}}$ and $\widetilde{\boldsymbol{\Theta}}$ parameter estimates from 300 simulations. True spectral parameters are indicated by the continuous lines.

TABLE I
QUALITY OF THE ESTIMATED STANDARD ERRORS

	θ^{loc}	θ^{or}	$\widetilde{\sigma}_k^2$
SNR 1:1			
q_{GLS}	0.86	0.91	0.67
q_{ML}	0.99	1.02	1.01
SNR 1:5			
q_{GLS}	0.88	0.91	0.65
q_{ML}	1.01	1.14	1.01

	$[\widetilde{\boldsymbol{\Psi}}]_{11}$	$\text{Re}[\widetilde{\boldsymbol{\Psi}}]_{12}$	$\text{Im}[\widetilde{\boldsymbol{\Psi}}]_{12}$	$[\widetilde{\boldsymbol{\Psi}}]_{22}$
SNR 1:1				
q_{GLS}	0.89	0.73	0.66	0.88
q_{ML}	0.99	1.01	0.85	0.99
SNR 1:5				
q_{GLS}	0.87	0.89	0.31	0.86
q_{ML}	0.99	1.01	0.85	0.99

The quality of the estimated standard errors (mean estimated standard error/simulation standard error), averaged for location and orientation parameters, and averaged across frequencies for $\widetilde{\boldsymbol{\sigma}}_k^2$ and $\widetilde{\boldsymbol{\Psi}}$ matrix elements. values close to one indicate good quality, values larger than 1 indicate overestimation of the true parameter standard errors, values smaller than 1 indicate underestimation of the true parameter standard errors.

TABLE II
RMSEs

	θ^{loc}	θ^{or}	$\widetilde{\sigma}_k^2$
SNR 1:1			
GLS	0.003	0.012	6.110
ML	0.002	0.003	0.223
SNR 1:5			
GLS	0.008	0.013	30.660
ML	0.006	0.008	1.118

	$[\widetilde{\boldsymbol{\Psi}}]_{11}$	$\text{Re}[\widetilde{\boldsymbol{\Psi}}]_{12}$	$\text{Im}[\widetilde{\boldsymbol{\Psi}}]_{12}$	$[\widetilde{\boldsymbol{\Psi}}]_{22}$
SNR 1:1				
GLS	0.223	0.156	0.218	0.533
ML	0.062	0.060	0.063	0.127
SNR 1:5				
GLS	0.234	0.179	0.233	0.545
ML	0.138	0.112	0.109	0.209

RMSEs estimated from the deviations of estimators around their true values.

timators, since their bias was small, their estimated standard errors had good quality, and they were approximately normally distributed. To get an impression the confidence intervals, the coverage rates are given in Table III. The confidence regions for the ML source location and orientation parameters θ are reliable (their coverage rate approximates the nominal rate of 95%). For the ML $\widetilde{\boldsymbol{\Psi}}$ confidence intervals the coverage rates are also close to the nominal rate of 95%. Coverage rates for ML $\widetilde{\boldsymbol{\Theta}}$ deviate substantially, especially for bad SNR. This is due to relatively small standard errors. These parameters, however, are to be considered as nuisance parameters, not of practical interest. Furthermore, they seem to have little effect on the other parameters. The coverage rates of the GLS θ confidence intervals were very close to the nominal level, but this was not the case for the $\widetilde{\boldsymbol{\Psi}}$ and $\widetilde{\boldsymbol{\Theta}}$ confidence intervals as should be expected from the magnitude of their bias and the underestimated parameter standard errors.

Because the GLS estimators did not perform well in the first set of simulations, we only examined the performance of the chi-square goodness of fit statistic with ML estimators. In 255 out of 300 simulations (87%), the two-dipole model was rejected by the chi-square test of size $\alpha = 0.05$, whereas the three-dipole model was selected in 279 out of the same 300 simulations (92%) which is close to the expected nominal rate of 95%. A Kolmogorov–Smirnov test on the chi-square variate did not indicate deviation from the appropriate chi-square distribution when the true, three-dipole model was used. These results indicate that the chi-square test statistic based on $F_{\mathrm{ML}}(\hat{\xi})$ can be helpful in determining the number of sources that are necessary, provided that the model is otherwise correctly specified.

D. Starting Values and Algorithm Convergence

Common problems with nonlinear optimization algorithms are, among others, obtaining good starting values and avoiding local minima. Even with the starting values near ξ_0 the ML method has a tendency to diverge to singular cross-spectral matrices $\widetilde{\boldsymbol{\Sigma}}_k(\xi)$. This could be circumvented by using ULS estimates as starting values in the ML iterations. Local minima could be avoided when multiple starting values were used. The GLS seemed to suffer a lot from local minima, and sometimes diverged to models with nonpositive definite amplitude cross-spectral matrices $\widetilde{\boldsymbol{\Psi}}_k$, especially with low SNR. In the simulations, when the maximum number of iterations was set to 500, on average about 300 major iterations (evaluation of the loss-function and its gradients) and two different starting values were necessary for convergence to an admissible solution $\hat{\xi}$.

IV. Application to MEG Data

As an illustration, we analyzed data obtained in a visual evoked field experiment from a 20-year-old subject of good health.

A. Data Collection

MEG data were obtained in a visual (right half field) stimulation experiment with the same checkerboard stimuli as used in [37]. A CTF Systems, Inc. gradiometer system with 151 gradiometers in a magnetically shielded room, recorded the MEG-signals. The subject sat silently while watching the stimuli. 392 trials were recorded.

B. Data Processing

An interval of 300 ms was selected, 82 ms prestimulus, and 211 ms poststimulus. The signals were sampled at a rate of 208.3 Hz, after low-pass filtering at 70 Hz. For each trial a prestimulus baseline of 48 ms was used to center the trial data. After removing the trial average from the signals, the signals were Fourier transformed and an estimate of the cross- spectral matrix was obtained at each frequency, as discussed following (5). A band of frequencies was chosen, in which the topographical contour plot of the sensors' signal power was approximately equal (correlations between 0.94 and 0.99). The coherence between sensors separated more than 9 cm is graphically displayed in Fig. 4(a). The plot suggests the presence of two coherent sources in the left and right temporal lobe. Therefore, a two-dipole model was fitted on the cross-spectral matrices of the four frequency components in the chosen frequency band (viz. 23.5, 26.9, 30.2, and 33.6 Hz.)

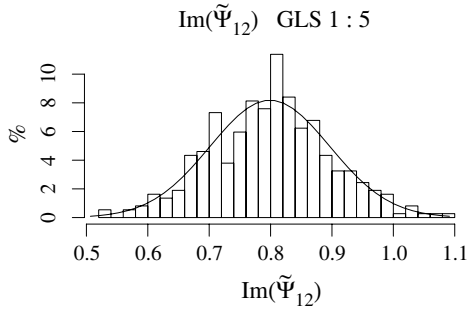

Fig. 3. Percentages of simulations in which $\widetilde{\boldsymbol{\Psi}}$ estimates were contained in the ranges indicated by the bars. The curved line indicates the percentage expected on the basis of the best fitting normal distribution. This figure gives the worst case deviance from normality. Even in this worst case the normal distribution seems a good approximation to the actual distribution.

TABLE III
Coverage Rates of the 95% Confidence Intervals

	θ	$\widetilde{\boldsymbol{\Psi}}$	$\widetilde{\sigma}_k^2$
SNR 1:1			
$\%_{GLS}$	92.12	14.30	0.00
$\%_{ML}$	95.71	93.79	71.42
SNR 1:5			
$\%_{GLS}$	84.50	51.10	0.00
$\%_{ML}$	96.26	94.59	43.67

Percentage of simulations of which 95% confidence intervals contained the true parameter values.

C. Results

The parameter estimates are presented in Table IV along with there estimated standard errors. The parameter standard errors of the two-dipole model were very large. We, therefore, subsequently fitted a three-dipole model. The estimated sources are located medially in the left and right hemisphere near the vertex. Two of the dipoles are located close to the sources in the two-dipole model. It is seen in Table IV that the standard errors obtained from the estimation procedure are still very large (although substantially smaller than for the two-dipole model). Also given in Table IV are the coherences between sources. These coherences were computed from the $\widetilde{\boldsymbol{\Psi}}$ estimates by $\hat{\rho}_{de}(k) = |(\widetilde{\boldsymbol{\Psi}}_k)_{de}|^2/((\widetilde{\boldsymbol{\Psi}}_k)_{dd}(\widetilde{\boldsymbol{\Psi}}_k)_{ee})^{1/2}$ [45]. The predicted coherence between sensors is displayed in Fig. 4(b) by the lines joining the sensors. Both the model and the data show high coherence between lateral temporal located sensors. For both models the chi-square test indicated lack of fit, which can be due to an incorrect head model, an incorrect number of sources or inadequacy of the dipole model for these sources,

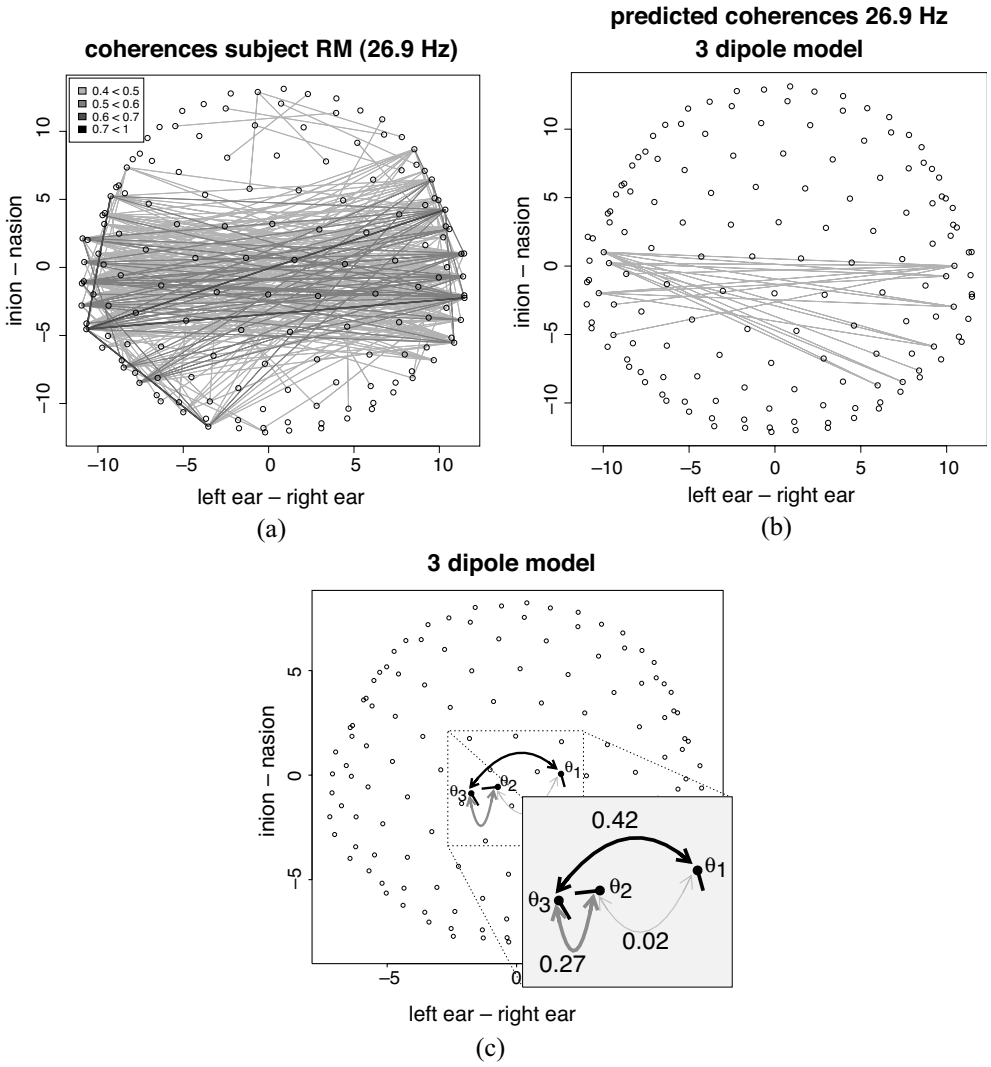

Fig. 4. (a) Coherence between sensors with distances greater than 9 cm. (b) Coherence between sensors predicted by the three-dipole model. (c) Three-dipole model—the arrows connecting sources indicate the coherence (averaged across the frequencies presented Table IV) between the source activity (see text for details).

or an incorrect noise model. In the current implementation, the noise cross-spectrum matrix is assumed to be proportional to the identity matrix which is an unrealistic assumption [46]. We, therefore, cannot draw definite conclusions about the number of sources from this test at present.

D. Conclusion

In Fig. 4(c), the dipole model is depicted, along with curves that indicate the coherence between sources. The dipole fit indicates that medial areas, likely in parietal cortical areas, are interacting. Since the sources are in close proximity of each other, the large standard errors might indicate that the underlying source is extended and is, therefore, not adequately modeled by dipole sources, but the pattern of coherence does not completely support this interpretation: One of these areas (θ_3) seems to be coherent with the other areas (θ_1 and θ_2), whereas the other areas lack coherency. In contrast, the sensor to sensor coherence estimates seem to indicate that areas much lower and more temporal are synchronizing activity. Based on the simulations, we should have more confidence in the dipole model than in the sensor coherence. However, the standard errors of the model parameters are high and the chi-square test indicates lack of fit. Furthermore incorrect assumptions about the noise cross-spectrum $\widetilde{\mathbf{\Theta}}_k$ also lead to biased estimates of $\widetilde{\mathbf{\Psi}}_k$ and subsequently over- or underestimated coherences between sources. The coherences should, therefore, be interpreted with caution. These somewhat disappointing results indicate that in order to make more definite inferences on these issues, the subtleties of the data have to be appreciated to a greater extent. The major source of lack of fit is presumably the incorrectness of the structure of the noise cross-spectrum $\widetilde{\mathbf{\Theta}}_k$. Earlier we indicated that this matrix might be more adequately modeled by a matrix $\widetilde{\sigma}_k^2\mathbf{U}$, where $\mathbf{U}$ is a known matrix. The random dipole model developed in [46] may provide an appropriate choice of $\mathbf{U}$. This model assumes that the background brain activity is generated by dipole like sources that are randomly located in the head, and have random orientations. This model predicts both the background EEG and the MEG quite well [46]. As the locations are random from time to time and trial to trial, they do not take part in the gain matrix

TABLE IV
THREE-DIPOLE MODEL FIT

	θ_x^{loc}	θ_y^{loc}	θ_z^{loc}	θ_x^{or}	θ_y^{or}	θ_z^{or}
dipole						
1	-1.9	2.0	4.7	0.76	-0.42	0.49
(s.e.)	(9.5)	(7.0)	(6.2)	(1.1)	(2.1)	(1.5)
2	-1.1	-1.9	4.8	0.66	0.63	0.40
(s.e.)	(7.6)	(8.0)	(5.9)	(1.4)	(1.9)	(1.6)
coh.	23.5	26.9	30.2	33.6	(Hz)	
$\hat{\rho}_{\theta_1 \leftrightarrow \theta_2}$	0.34	0.44	0.55	0.46		
dipole						
1	0.4	-2.9	4.7	-0.97	-0.22	-0.05
(s.e.)	(5.2)	(3.7)	(3.1)	(0.3)	(1.4)	(1.2)
2	-0.6	0.9	5.3	-0.13	0.98	-0.17
(s.e.)	(4.5)	(7.4)	(4.1)	(3.1)	(0.5)	(1.4)
3	-1.1	2.5	5.5	-0.90	-0.43	0.02
(s.e.)	(6.8)	(4.5)	(3.4)	(1.2)	(2.4)	(1.7)
coh.		23.5	26.9	30.2	33.6	(Hz)
$\hat{\rho}_{\theta_1 \leftrightarrow \theta_2}$		0.00	0.03	0.05	0.02	
$\hat{\rho}_{\theta_1 \leftrightarrow \theta_3}$		0.44	0.46	0.37	0.42	
$\hat{\rho}_{\theta_2 \leftrightarrow \theta_3}$		0.21	0.31	0.31	0.24	

The θ^{loc} parameter values are in centimeter units, θ^{or} parameters are direction cosines. $\hat{\rho}_{\theta_k \leftrightarrow \theta_m}$ is the estimated coherence between the sources θ_k and θ_m at the indicated frequencies.

$\mathbf{G}(\theta)$ which only models the sources that are consistent across time and trials. Incorporating such a model for the noise covariance seems vital for successful application of the method.

V. DISCUSSION

We conclude from the simulations that with the method presented, it is possible to obtain accurate estimates of sources together with measures of interaction of these sources. In theory, both GLS and ML should be adequate, however, the theory is based on large numbers of trials (L) and can only be an approximation to the finite number of trials situation of actual experiments. In our simulation with a moderate trial count $(L = 400)$, the GLS estimation procedure turned out to give biased estimators and poor quality standard error estimates, and as a consequence, poor quality confidence intervals. In contrast, the ML procedure provided (apart from the nuisance $\widetilde{\mathbf{\Theta}}$ estimators) nearly unbiased estimators, with adequate standard errors, and reliable confidence intervals.

In Fig. 1, the method was compared with conventional coherence analysis of the data in a typical simulation. Interpretation of the coherences displayed in the figure would probably lead to the conclusion that central sensorimotor areas are interacting with more lateral areas on the same side of the head, possibly with contralateral temporal areas as well. This interpretation misses the two interacting sources in the parietal areas (one in each hemisphere). This shows that in general it will be very difficult to discern the amount of coherence due to cortical interactions and the amount of coherence due to overlapping sensor sensitivities. Although an additional phase analysis would help to some extent, this still does not provide a clear picture of the location of the sources, and limits the analysis to out-of-phase sources. The proposed method on the other hand, does provide an interpretation both in terms of physiological interactions and anatomical origin.

A difficulty may be observed in two somewhat conflicting requirements of the method: On the one hand the time series segments have to be sufficiently large for the Fourier coefficients to become uncorrelated, while on the other hand the signals are assumed to be generated by relatively small number of dipole like sources. The number of sources incorporated in $\mathbf{G}(\theta)$ will not be very high, since the cross-spectrum estimates are based on the periodograms of the signals averaged across trials. Therefore, the power contribution of sources not consistently present at each trial will be diminished and smoothed out as theorized in [46]. Selection of a relatively small window size, which may be necessary to ensure that only a few dipoles are active, will introduce error in the approximations of (7) and (8) which in turn introduces some inefficiency in the estimates. We plan to investigate this error of approximation in a future manuscript.

We used the chi-square goodness of fit approach to assess the appropriateness of the model in the simulations. In the simulations, this chi-square test turned out to be quite helpful in determining how many sources should be included in the ECD model. In real data, however, the chi-square test is also sensitive to a wrong model for the noise cross-spectrum or a wrong head model. A better noise model than $\widetilde{\mathbf{\Theta}}_k = \widetilde{\sigma}_k^2 \mathbf{I}$ and realistic head models may alleviate this problem. Possible candidates for $\widetilde{\mathbf{\Theta}}$ that accommodate noise correlations between neighboring sensors are given in [46], as discussed previously, and [47] and [48]. However, some model misspecification will be always present because of necessary approximations to reality. The chi-square will then reject a model because of the approximation and not because of an inadequate number of sources. Therefore, alternative measures of fit have to be considered [42].

It was noted in the simulations that convergence can be rather slow, especially with large numbers of sensors (≥ 61). The rate of convergence of course depends in part on the starting values for the parameters, the number of frequencies included, and the number of model parameters. Good starting values can quicken convergence considerably, but are not generally easy to find. We obtained starting values from ULS estimates. This did improve the convergence rate of the ML procedure considerably and, as an additional advantage, prevented divergence to improper solutions (i.e., nonpositive definite cross-spectral matrices). Also, we first obtained estimates at each frequency independently, and then used the frequency average of these parameter estimates as starting values for the simultaneous analysis of all frequencies of interest. Further strategies to find starting values may be found, which may include the following: First, a small subset of sensors is selected covering the regions in which sources are expected (e.g., based on previous research). Then, from this sensor-subset

ULS estimates are obtained and used as starting values in a full ML optimization procedure with all sensors included. It is not always necessary to include all sensors; it is important to make a selection of those sensors that provide the most information about the sources that are expected to be found. Methods to make such optimal selections are given in [49].

The frequency domain approach to ECD modeling of MEG/EEG has been pursued earlier in [23]–[25] and [50]. The current work may be seen as an extension of the methods presented there. In [23]–[25], source amplitudes were assumed deterministic across trials. In [50], variations in magnitude and latency of the source amplitude waveforms were allowed for different experimental conditions. No attempts were made to estimate source covariation in these papers. In this paper, source amplitudes were assumed to be partly stochastic, and this stochastic nature was used to estimate their covariation across trials. It may be noted that the method presented here is close to what is known as "stochastic ML" (SML) in the signal processing literature [51].[3]

The method of this paper may be extended by further parameterizing $\widetilde{\mathbf{\Psi}}$ to include structural regressions (linear or nonlinear) of activities from one source on another (i.e., estimates of transfer functions). The method may also be extended to include a model for the trial average signal. Furthermore, the method may be extendable to the time domain, where the assumption of stationarity is unnecessary. We are currently working on such extensions. Computational simplifications may be found along the lines that are given in the discussion of SML in [51], and will be pursued in future work.

Although the method was tested with MEG, it is not expected that the results will be drastically altered when using EEG. Also, in principle MEG and EEG could be combined in the analysis, thus profiting from their specific advantages [38].

ACKNOWLEDGMENT

The authors would like to thank C. V. Dolan for his helpful comments. They would also like to thank the anonymous reviewers for their effort and valuable comments.

REFERENCES

[1] M. I. Posner and M. E. Raichle, "Networks of attention," in *Images of Mind*, M. I. Posner, Ed. San Francisco, CA: Freeman, 1994, pp. 153–179.

[2] K. J. Friston and C. Büchel, "Attentional modulation of effective connectivity from V2 to V5/MT in humans," *Neuroimage*, vol. 14, pp. 1353–1360, 2001.

[3] J. H. Houtveen, B. Bermond, and M. R. Elton, "Alexithymia: a disruption in a cortical network? An EEG power and coherence analysis," *J. Psychophysiol.*, vol. 11, pp. 147–157, 1997.

[4] L. I. Aftanas, A. A. Varlamov, S. V. Pavlov, V. P. Makhnev, and N. V. Reva, "Affective picture processing: event-related synchronization within individually defined human theta band is modulated by valence dimension," *Neurosci. Lett.*, vol. 303, pp. 115–118, 2001.

[5] Z. J. Koles and P. Flor-Henry, "The effect of brain function on coherence patterns in the bipolar EEG," *Int. J. Psychophysiol.*, vol. 5, pp. 63–71, 1987.

[3]We thank an anonymous reviewer for bringing this to our attention.

[6] W. Singer, "Synchronization of neural responses as a putative binding mechanism," in *Handbook of Brain Research and Neural Networks*, M. A. Arbib, Ed. Cambridge, MA: M.I.T. Press, 1995, pp. 960–964.

[7] J. I. Nelson, "Binding in the visual system," in *Handbook of Brain Research and Neural Networks*, M. A. Arbib, Ed. Cambridge, MA: M.I.T. Press, 1995, pp. 157–159.

[8] P. L. Nunez, R. B. Silberstein, P. J. Caduch, R. S. Wijesinghe, A. F. Westdorp, and R. Srinivasan, "A theoretical and experimental study of high resolution EEG based on surface laplacian and cortical imaging," *Electroencephalogr. Clin. Neurophysiol.*, vol. 90, pp. 40–57, 1994.

[9] M. Hämäläinen, R. Hari, R. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, "Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain," *Rev. Modern Phys.*, vol. 65, pp. 413–497, 1993.

[10] K. J. Friston, C. Buechel, G. R. Fink, J. Morris, E. Rolls, and R. J. Dolan, "Psychophysiological and modulatory interactions in neuroimaging," *Neuroimage*, vol. 6, pp. 218–229, 1997.

[11] M. S. Gonçalves, D. A. Hall, I. S. Johsrude, and M. P. Haggard, "Can meaningful effective connectivities be obtained between auditory cortical regions," *Neuroimage*, vol. 14, pp. 1353–1360, 2001.

[12] A. R. McIntosh, C. L. Grady, L. G. Ungerleider, J. V. Haxby, S. I. Rapoport, and B. Horwitz, "Network analysis of cortical visual pathways mapped with PET," *J. Neurosci.*, vol. 14, no. 2, pp. 655–666, 1994.

[13] D. M. Tucker, D. L. Roth, and T. B. Bair, "Functional connections among cortical regions: topography of EEG coherence," *Electroencephalogr. Clin. Neurophysiol.*, vol. 63, pp. 242–250, 1986.

[14] C. Andrew and G. Pfurtscheller, "Event-related coherence as a tool for studying dynamic interaction of brain regions," *Electroencephalogr. Clin. Neurophysiol.*, vol. 98, pp. 144–148, 1996.

[15] M. Kaminski, K. Blinowska, and W. Szelenberger, "Topographic analysis of coherence and propagation of EEG activity during sleep and wakefulness," *Electroencephalogr. Clin. Neurophysiol.*, vol. 102, pp. 216–227, 1997.

[16] P. C. M. Molenaar, "Dynamic factor analysis of psychophysiological signals," in *Advances in Psychophysiology*, J. R. Jennings, P. Ackles, and M. G. H. Coles, Eds. London, U.K.: Kingsley, 1993, vol. 5, Advances in Psychophysiology, pp. 229–302.

[17] G. Pfurtscheller and F. H. Lopez da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, 1999.

[18] R. Srinivasan, P. L. Nunez, and R. B. Silberstein, "Spatial filtering and neocortical dynamics: estimates of EEG coherence," *IEEE Trans. Biomed. Eng.*, vol. 45, pp. 814–826, July 1998.

[19] P. Rappelsberger, "The reference problem and mapping of coherence: a simulation study," *Brain Topogr.*, pp. 63–72, 1989.

[20] R. D. Katznelson, "EEG recording, electrode placement, and aspects of generator localization," in *Electric Fields of the Brain: The Neurophysics of EEG*, P. L. Nunez, Ed. New York: Oxford Univ. Press, 1981, pp. 176–213.

[21] I. A. Cook, R. O'Hara, S. H. J. Uijtdehaage, M. Mandelkern, and A. F. Leuchter, "Assessing the accuracy of topographic EEG mapping for determining local brain function," *Electroencephalogr. Clin. Neurophysiol.*, vol. 107, no. 6, pp. 408–414, 1998.

[22] J. Gross, J. Kujala, M. Hämäläinen, L. Timmermann, A. Schnitzler, and R. Salmelin, "Dynamic imaging of coherent sources: studying neural interactions in the human brain," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 2, pp. 694–699, 2001.

[23] J. Raz, B. Turetsky, and G. Fein, "Frequency-domain estimation of the parameters of human brain electrical dipoles," *J. Amer. Statist. Assoc.*, vol. 87, no. 417, pp. 69–77, 1992.

[24] J. Raz, C. A. Biggins, B. Turetsky, and G. Fein, "Frequency-domain dipole localization—extensions of the method and applications to auditory and visual-evoked potentials," *IEEE Trans. Biomed. Eng.*, vol. 40, pp. 909–918, Sept. 1995.

[25] B. Lütkenhöner, "Frequency-domain localization of intracerebral dipolar sources," *Electroencephalogr. Clin. Neurophysiol.*, vol. 82, no. 2, pp. 112–118, 1992.

[26] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus: With Applications in Statistics and Econometrics*, ser. Probability and Statistics. Chichester: Wiley, 1999.

[27] D. R. Brillinger, *Time Series: Data Analysis and Theory*, ser. Decision Processes, New York: Holt, Reinhart and Winston, 1975.

[28] K. G. Jöreskog, "Analysis of covariance structures," *Scand. J. Statist.*, vol. 8, pp. 65–92, 1981.

[29] K.-H.Ke-Hai Yuan and P. M.Peter M. Bentler, "Mean and covariance structure analysis: theoretical and practical improvements," *J. Amer. Statist. Assoc.*, vol. 92, no. 438, pp. 767–774, 1997.

[30] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[31] M. W. Browne, "Generalized least squares estimators in the analysis of covariance structures," *S. African Statist. J.*, vol. 8, pp. 1–24, 1974.

[32] ——, "Asymptotically distribution-free methods for the analysis of covariance structures," *Br. J. Math. Statist. Psych.*, vol. 37, pp. 62–83, 1984.

[33] S. D. Silvey, *Statistical Inference*. Harmondsworth, U.K.: Penguin, 1970.

[34] A. Shapiro and M. W. Browne, "Analysis of covariance structures under elliptical distributions," *J. Amer. Statist. Assoc.*, vol. 82, no. 400, pp. 1092–1097, 1987.

[35] H. M. Huizenga and P. C. M. Molenaar, "Estimating and testing the sources of evoked potentials in the brain," *Multivariate Behavioral Res.*, vol. 28, pp. 237–262, 1994.

[36] J. C. de Munck, A. de Jongh, and B. W. van Dijk, "The localization of spontaneous brain activity: an efficient way to analyze large sata sets," *IEEE Trans. Biomed. Eng.*, vol. 48, pp. 1221–1228, Nov 2001.

[37] J. L. Kenemans, J. M. P. Baas, G. R. Mangun, M. Lijffijt, and M. N. Verbaten, "On the processing of spatial frequencies as revealed by evoked-potential source modeling," *Clin. Neurophysiol.*, vol. 111, no. 6, pp. 1113–1123, 2000.

[38] H. M. Huizenga, T. L. Van Zuijen, D. J. Heslenfeld, and P. C. M. Molenaar, "Simultaneous MEG and EEG source analysis," *Phys. Med. Biol.*, vol. 46, no. 7, pp. 1737–1751, 2001.

[39] J. Aitchinson and S. D. Silvey, "Maximum likelihood estimation of parameters subject to restraints," *Ann. Math. Statist.*, vol. 29, pp. 813–828, 1958.

[40] M. W. Browne and S. H. C. du Toit, "Automated fitting of nonstandard models," *Multivariate Behavioral Res.*, vol. 27, pp. 269–300, 1992.

[41] D. F. Morrison, *Multivariate Statistical Methods*, 2d ed. New York: McGraw Hill, 1989.

[42] L. J. Waldorp, H. M. Huizenga, R. P. P. P. Grasman, K. B. E. Böcker, J. C. de Munck, and P. C. M. Molenaar, "Model selection in electromagnetic source analysis with an application to VEFs," *IEEE Trans. Biomed. Eng.*, vol. 49, pp. 1121–1129, Oct. 2002.

[43] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright, *User's Guide for NPSOL (Version 4.0): A FORTRAN Package for Nonlinear Programming*, 4.0 ed. Stanford, CA: SOL, 1998.

[44] C. V. Dolan and P. C. M. Molenaar, "A comparison of four methods of calculating standard errors of maximum likelihood estimates in the analysis of covariance structure," *Br. J. Math. Statist. Psych.*, vol. 44, pp. 359–369, 1991.

[45] L. H. Koopmans, *The Spectral Analysis of Time Series*, 2nd ed. San Diego, CA: Academic, 1995, vol. 22, Probability and Mathematical Statistics.

[46] J. C. de Munck, P. C. M. Vijn, and F. H. Lopes da Silva, "A random dipole model for spontaneous brain activity," *IEEE Trans. Biomed. Eng.*, vol. 39, pp. 986–990, Aug. 1992.

[47] J. C. De Munck, H. M. Huizenga, L. J. Waldorp, and R. M. Heethaar, "Estimating stationary dipoles from MEG/EEG data contaminated with spatially and temporally correlated background noise," *IEEE Trans. Signal Processing*, vol. 50, pp. 1565–1572, July 2002.

[48] L. J. Waldorp, H. M. Huizenga, C. V. Dolan, and P. C. M. Molenaar, "Estimated generalized least squares electromagnetic source analysis based on a parametric noise covariance model," *IEEE Trans. Biomed. Eng.*, vol. 48, pp. 737–741, June 2001.

[49] H. M. Huizenga and P. C. M. Molenaar, "Optimal measurement conditions for spatiotemporal EEG/MEG source analysis," *Psychometrika*, vol. 67, no. 2, pp. 299–313, 2002.

[50] J. Raz, V. Cardenas, and D. Fletcher, "Frequency-domain estimation of covariate effects in multichannel brain evoked-potential data," *Biometrics*, vol. 51, no. 2, pp. 448–460, 1995.

[51] H. Krim and M. Viberg, "Two decades of array signal processing research. The parametric approach," *IEEE Signal Processing Mag.*, vol. 13, pp. 67–94, July 1996.

Raoul P. P. P. Grasman (S'00–A'03) was born in 1973. He received a degree in artificial intelligence (*cum laude*) and the M.S. in experimental psychology (*cum laude*) from the University of Amsterdam, Amsterdam, The Netherlands, in 1997 and 1998, respectively. He is currently working towards the Ph.D. degree at the University of Amsterdam.

His research interests concern the methodology of cognitive neuroscience and experimental psychology research.

Hilde M. Huizenga was born in 1965. She received the M.A. degree in psychology from the University of Groningen, Groningen, The Netherlands, in 1990 and the Ph.D. degree (*cum laude*) in psychology from the University of Amsterdam, Amsterdam, The Netherlands, in 1995.

From 1996 to 2001, she was a Postdoctoral Fellow at the University of Amsterdam where she is currently a Senior University Lecturer. Her main research interest is statistical analysis of neuro-scientifc data; in particular, nonlinear regression and covariance structure analysis of EEG/MEG sources and their interactions.

Lourens J. Waldorp (S'00–A'03) was born in 1971. He received the M.S. degree in methodological psychology from the University of Amsterdam, Amsterdam, The Netherlands, in 1998. He is currently working towards the Ph.D. degree at the University of Amsterdam.

His research interests include statistical analyses in psychophysiological experiments and signal processing.

Koen B. E. Böcker received the M.S. degree in physiological psychology and the Ph.D. degree from Tilburg University, Tilburg, The Netherlands, in 1989 and 1994, respectively.

He is currently a Postdoctoral Researcher with the Department of Psychopharmacology at Utrecht University, Utrecht, The Netherlands, working on a cooperative project, together with the other authors, on statistical EEG/MEG source analysis. Since 1989, he has served multiple research positions at the universities of Tilburg, Eindhoven (Technical University), and Utrecht in The Netherlands, and Liège and Brussels (Free University) in Belgium. His research interests include the study of perception (visual, spoken language) and cognition (imagery, attention, and emotion) and the application of source analysis in cognitive neuroscience.

Peter C. M. Molenaar was born in 1946. His Ph.D. dissertation was on multidimensional signal analysis.

He is Research Director of several programs and a Department Head at the University of Amsterdam, Amsterdam, The Netherlands. His current research interests include signal analysis and applied nonlinear dynamics.