

Evaluation of EEG localization methods using realistic simulations of interictal spikes

C. Grova,^{a,*} J. Daunizeau,^{b,d} J.-M. Lina,^{b,c,d} C.G. Bénar,^a H. Benali,^{b,d} and J. Gotman^a

^aMontreal Neurological Institute, McGill University, EEG Department, Room 009d, 3801 University Street, Montreal, Quebec, Canada H3A 2B4

^bCentre de Recherche en Mathématiques, Université de Montréal, Montreal, Canada

^cDépartement de Génie Électrique, Ecole de Technologie Supérieure, Montreal, Canada

^dUMR 678 INSERM/UPMC, Laboratoire d'Imagerie Fonctionnelle, CHU Pitié Salpêtrière, Paris, France

Received 7 December 2004; revised 6 July 2005; accepted 23 August 2005
Available online 3 November 2005

Performing an accurate localization of sources of interictal spikes from EEG scalp measurements is of particular interest during the presurgical investigation of epilepsy. The purpose of this paper is to study the ability of six distributed source localization methods to recover extended sources of activated cortex. Due to the frequent lack of a gold standard to evaluate source localization methods, our evaluation was performed in a controlled environment using realistic simulations of EEG interictal spikes, involving several anatomical locations with several spatial extents. Simulated data were corrupted by physiological EEG noise. Simulations involving pairs of sources with the same amplitude were also studied. In addition to standard validation criteria (e.g., geodesic distance or mean square error), we proposed an original criterion dedicated to assess detection accuracy, based on receiver operating characteristic (ROC) analysis. Six source localization methods were evaluated: the minimum norm, the minimum norm weighted by multivariate source prelocalization (MSP), cortical LORETA with or without additional minimum norm regularization, and two derivations of the maximum entropy on the mean (MEM) approach. Results showed that LORETA-based and MEM-based methods were able to accurately recover sources of different spatial extents, with the exception of sources in temporo-mesial and fronto-mesial regions. Several spurious sources were generated by those methods, however, whereas methods using the MSP always located very accurately the maximum of activity but not its spatial extent. These findings suggest that one should always take into account the results from different localization methods when analyzing real interictal spikes.

© 2005 Elsevier Inc. All rights reserved.

Keywords: EEG; Source localization; Validation; ROC; Regularization techniques; Interictal spikes

Introduction

Presurgical evaluation of patients with epilepsy is a multiple step process, whose main goal is the identification of the region of the brain that must be resected in order to abolish epileptic seizures. Such an area is thought to be organized as an epileptogenic network (Chauvel et al., 1996; Spencer, 2002), whose identification remains a challenging issue. Electroencephalography (EEG) plays a crucial role as it is the only technique that provides information specific to the epileptic state, during seizures (the “ictal” state) or between seizures (the “interictal” state). Interictal spikes are transient events, characteristic of epilepsy, that occur between seizures. They are generated by the brain without any clinical signs, thus allowing multimodal imaging studies. Studying the underlying mechanisms and anatomical areas involved in the generation and the propagation of interictal spikes constitutes a key issue toward a better understanding of epileptic disorders (Luders and Awad, 1992). Scalp EEG recordings have an excellent temporal resolution (~1 ms), but its visual interpretation provides only imprecise localization, indicating at best which lobe was involved during the epileptic discharge. EEG source localization methods may help to determine the regions of the brain where the spikes are generated (Ebersole, 1997; Merlet and Gotman, 1999; Michel et al., 2004b). Other imaging modalities allow the exploration of the interictal state by studying blood perfusion (Devous et al., 1998), glucose metabolism (Chassoux et al., 2004), or the hemodynamic response that correlates with EEG activity (Al-Asmi et al., 2003). Even though those techniques provide a better spatial resolution than scalp EEG, only EEG (Ebersole, 1997; Merlet and Gotman, 1999; Michel et al., 2004b) and magnetoencephalography (Baumgartner, 2004) provide information directly linked to the generation of the spikes. Performing an accurate localization of EEG sources of interictal spikes is thus of particular interest to better understand their generation and propagation.

Inferring the source location within the brain from a signal acquired on the scalp, i.e., the EEG inverse problem, is an ill-posed problem since there is an infinite number of source configurations that can produce the exact same potential at the surface of the head.

* Corresponding author. Fax: +1 514 398 8106.

E-mail address: christophe.grova@mail.mcgill.ca (C. Grova).

Available online on ScienceDirect (www.sciencedirect.com).

Additional constraints should then be added to obtain a unique solution. Two types of approaches have been proposed (Baillet et al., 2001; Michel et al., 2004b). (1) The “equivalent current dipole” methods assume that the potentials are generated by a few dipolar sources (Scherg and Von Cramon, 1986; Koles, 1998). (2) The “distributed source” methods assume that surface potentials are generated by a large number of dipolar sources distributed within the brain or on the cortical surface (Dale and Sereno, 1993). Dipole scanning methods, assessing the presence of an equivalent current dipole on each point of a 3D grid, constitute an intermediate alternative to estimate 3D map of activation indices (Mosher and Leahy, 1999; Sekihara et al., 2001; Grave de Peralta Menendez et al., 2001; Bénar et al., 2005).

Whereas distributed source methods seem appropriate to estimate spatially extended sources, they aim at localizing neuronal activation on a large number of sources (around 10,000) given few sensors (around 100 or less). The problem is then highly under-determined and requires additional constraints in the form of a regularization scheme (Tikhonov and Arsenin, 1977). Some anatomical constraints were proposed such as localizing the sources on the cortical surface (Dale and Sereno, 1993), as well as some more “mathematical” constraints such as choosing the minimum energy solution (Hamalainen and Ilmoniemi, 1994) or the maximum of spatial smoothness solution (Pascual-Marqui et al., 1994). Statistical frameworks, based on Bayesian inference (Schmidt et al., 1999; Trujillo-Barreto et al., 2004) or entropy (Clarke and Janday, 1989; Amblard et al., 2004), were also proposed to provide a flexible way of introducing prior information. Prior information may consist of anatomical information (Trujillo-Barreto et al., 2004) and/or functional information (Liu et al., 1998; Mattout et al., 2005).

In the context of interictal spike localization, equivalent current dipole methods were proved to be useful to localize the “center of mass” of brain activity (Merlet and Gotman, 1999; Stephen et al., 2003; Kobayashi et al., 2003). Intracerebral recordings showed that activated brain areas able to generate spikes on the scalp are rarely focal on the cortex, whereas focal deep intracerebral sources are almost never detected on surface EEG (Alarcon et al., 1994; Merlet and Gotman, 1999). A minimum area of cortex needs to be involved in order for the spike to be visible on the scalp, at least several cm^2 . The figure of 1 in^2 (approximately $2.5 \times 2.5 \text{ cm}^2$) has been proposed by Cooper et al. (1965), who performed measurements on a wet skull. Ebersole (1997) confirmed that this figure of around 6 cm^2 , although determined in a simplistic way, seems a good estimate of the minimum area required to see the spike on the scalp. Such value of 6 cm^2 should be considered as a low bound. Indeed, very widespread spikes involving most of one lobe are not rare.

The purpose of this paper was the validation of six distributed source localization methods in the clinical context of EEG interictal spikes localization. Our objective was to assess their detection accuracy when extended cortical areas were activated. Due to the frequent lack of a gold standard in the context of EEG source localization, we performed our evaluation in a fully controlled environment using realistic simulations of EEG interictal spikes, involving several anatomical locations with several spatial extents. All simulated data were corrupted by averaged physiological EEG noise. Two levels of signal-to-noise ratios as well as simulations involving pairs of sources with the same amplitude were also studied. In addition to usual validation criteria, such as the geodesic distance to the global maximum of activity or mean square errors on the reconstructed activity (Im et al., 2003),

we proposed an original criterion to assess detection accuracy. This validation metric was mainly based on receiver operating characteristic (ROC) curves (Metz, 1986), adapted to the context of distributed source localization methods, in order to better estimate the false positive rate far and within a close neighborhood of the simulated source. Six source localization methods were evaluated and compared, namely: the minimum norm (MN) (Hamalainen and Ilmoniemi, 1994), the minimum norm weighted by the multivariate source prelocalization (MSP) (Mattout et al., 2005), cortical LORETA with (David and Garnero, 2002) or without (Pascual-Marqui et al., 1994) additional regularization using a minimum norm term and two derivations of the maximum entropy on the mean (MEM) approach (Amblard et al., 2004).

Distributed source localization methods and regularization techniques

Distributed source localization methods assume that the anatomical substrate of brain activity is known, i.e., that the main generators of EEG potentials are the large pyramidal neurons of layer IV, which are oriented perpendicular to the surface of the cortex (Speckmann et al., 2004). When synchronously activated, each macro-column of those neurons behaves as an electrical dipole, observed from a distance. It is therefore reasonable to model the sources of EEG potentials as a large number of dipolar sources distributed on the cortex, which may be represented by the white matter/gray matter interface, the orientation of each dipole being fixed as the normal to the cortical surface. Under those assumptions, the relationship between source amplitudes and scalp potentials is expressed by the following linear model (Dale and Sereno, 1993):

$$\mathbf{M} = \mathbf{G} \cdot \mathbf{J} + \mathbf{E} \quad (1)$$

where \mathbf{M} is a $n \times t$ matrix of the EEG signal measured at n electrodes and t time samples. \mathbf{J} is the $p \times t$ unknown matrix of amplitudes of the p dipoles along the time. \mathbf{G} indicates the $n \times p$ forward operator. For each dipole of unit amplitude oriented along the normal to the cortical surface, the corresponding “forward field” (i.e., the corresponding column of \mathbf{G}) estimates the electric potentials generated at the n electrodes. Data are corrupted by an additive measurement noise \mathbf{E} ($n \times t$ matrix). Distributed source methods aim at estimating the amplitude of each source of the model $\hat{\mathbf{J}}$, so that the computed potentials $\mathbf{G} \cdot \hat{\mathbf{J}}$ are a best fit to the observed data \mathbf{M} . The least square solution $\hat{\mathbf{J}}_{\text{LS}}$ is then obtained by finding the best fit solution in the sense of a L2 norm:

$$\hat{\mathbf{J}}_{\text{LS}} = \operatorname{argmin}_{\mathbf{J}} \|\mathbf{M} - \mathbf{G} \cdot \mathbf{J}\|^2, \text{ i.e., } \hat{\mathbf{J}}_{\text{LS}} = (\mathbf{G}^T \cdot \mathbf{G})^+ \cdot \mathbf{G}^T \cdot \mathbf{M} \quad (2)$$

where, $+$ indicates the Moore–Penrose pseudo-inverse. However, as \mathbf{G} expresses the contribution of thousands of sources to very few sensors, it is ill-conditioned. Since error may arise from the forward problem and as data are corrupted by noise, the least square solution is generally not appropriate for solving the inverse problem, unless we add extra information. Such information usually represents some prior knowledge we may have of the current distribution \mathbf{J} . In this study, we used two probabilistic frameworks particularly suitable to incorporate prior knowledge in the EEG inverse problem: the Bayesian framework (Baillet et al., 2001) and the maximum entropy on the mean (MEM) (Amblard et al., 2004). Within the Bayesian framework (see Appendix A), prior information on the variance of the sources $\operatorname{Var}[\mathbf{J}]$ is

generally introduced in the solution by choosing $\text{Var}[\mathbf{J}]$ proportional to $(\mathbf{W} \cdot \mathbf{W}^T)^{-1}$ and notably by setting the weighting $p \times p$ matrix \mathbf{W} . The amount of regularization is tuned with an hyperparameter α (see Eq. (A.1)). Although no explicit optimum value can be found for α , many heuristic methods have been proposed to balance this trade-off between the fit to the data and the amount of prior information taken into account: the L curve (Hansen, 1992; Gorodnitsky et al., 1995), cross-validation (Bjorkstrom, 2000), the discrepancy principle (Morozov, 1968; Fuchs et al., 1999). On the other hand, the maximum entropy on the mean (MEM) principle is another regularization technique (see Appendix B), which aims at estimating, from the available data, the solution that makes the least assumption regarding missing information (Clarke and Janday, 1989; Amblard et al., 2004). Prior information regarding the distribution of \mathbf{J} is introduced by the definition of a reference distribution $d\mu$. The six localization methods to be evaluated are particular implementations of both those general frameworks.

Implementation of the source localization methods

In this section, we first briefly present the multivariate source prelocalization (MSP), a preprocessing method used to extract some *prior* information “from the data”, in an empirical Bayesian point of view. Such information will be used for both the weighted minimum norm solution (cf. definition of \mathbf{W}) and the initialization of the MEM (cf. definition of the reference distribution $d\mu$). The six localization methods to be evaluated are then presented.

Preprocessing to better condition the inverse problem

The multivariate source prelocalization (MSP) (Mattout et al., 2005) method was first used to generate activation probability map (APM) by estimating for each source of the model a probability-like coefficient of activation over a time window where the signal is assumed to be stationary. This was done by estimating functionally informed basis functions (fIBF) obtained from the singular value decomposition of $\hat{\mathbf{G}}$, the lead field matrix \mathbf{G} normalized by the norm of each column. A measure of correlation between those fIBF and the data \mathbf{M} was then used to estimate the contribution of each source of the model to the data: $\text{APM}(i)$ ($i \in \langle 1, p \rangle$). A coefficient $\text{APM}(i)$ close to 1 indicates that this source is very likely to contribute to the data. For all our simulated data, MSP was applied on a stationary time window of 10 time samples (see Simulation of the EEG interictal spike).

Localization methods to be evaluated

- The minimum norm (MN) method (Hamalainen and Ilmoniemi, 1994): the MN solution, obtained when \mathbf{W} is set to identity \mathbf{I}_p in Eq. (A.2), assumes that the solution of the inverse problem should be that providing the minimum energy of the current distribution.
- The MSP weighted minimum norm (WMN) method (Mattout et al., 2005): using the MSP results, we chose \mathbf{W} in Eq. (A.2) to be diagonal with $\mathbf{W}(i, i) = 1/\text{APM}(i)$ for each source $i \in \langle 1, p \rangle$. This WMN solution aims at estimating a minimum energy solution but relaxes the constraint on the sources that are more likely to be activated according to the MSP, allowing them to have a higher activity.

- LORETA-like methods: defining $\mathbf{W} = \Delta$ in Eq. (A.2), where Δ corresponds to the discrete Laplacian computed on the cortical mesh, we obtain a spatially smooth solution equivalent to the one provided by the low resolution electromagnetic tomography (LORETA) method (Pascual-Marqui et al., 1994). Two methods based on this assumption of spatial smoothness were actually evaluated:

- LORETA1: this method will refer directly to this LORETA-like approach proposed within the Bayesian framework (cf. $\mathbf{W} = \Delta$).
- LORETA2: some computational instabilities may occur when using the constraint $\mathbf{W} = \Delta$ computed on a 3D cortical mesh (David and Garnero, 2002). Additional regularization using a minimum norm term was therefore used by initializing the weighting matrix as follows: $\mathbf{W}^T \cdot \mathbf{W} = \Delta^T \cdot \Delta + \alpha_{\text{MN}} \mathbf{I}_p$. The hyperparameter α_{MN} was empirically set to $0.0001 \cdot \frac{\text{tr}(\Delta^T \cdot \Delta)}{\text{tr}(\mathbf{I}_p)}$ (tr refers to the trace of the matrix).
- MEM-like methods: in order to incorporate prior information in the MEM approach by the initialization of $d\mu$, we considered that brain activity may be described by K cortical parcels showing a homogeneous activation state. We first used a data-driven cortex parcelling method recently proposed by Daunizeau et al. (2004) in order to parcel the cortical surface into a set of spatially connected and functionally coherent components, exploiting results from the MSP. Such parcelling was obtained by first selecting K seed points as non-connected local maxima of the APM map followed by a region-growing approach in order to generate functionally coherent cortical parcels. The number of required parcels K was set a priori. Here, we used $K = 86$ parcels (mean surface of each parcel $21.6 (\pm 7.02) \text{ cm}^2$).

Each cortical parcel k was characterized by an activation state S_k (0 or 1), described by a probability of being active: $\text{Prob}(S_k = 1) = \alpha_k$ (Amblard et al., 2004). Assuming that the K parcels are mutually independent, prior information for the MEM was introduced by the reference distribution $d\mu$ defined by the following product of mixtures:

$$d\mu(\mathbf{j}) = \prod_{k=1}^K [(1 - \alpha_k)\delta(\mathbf{j}_k) + \alpha_k \mathcal{N}(\mu_k, \Sigma_k)(\mathbf{j}_k)] d\mathbf{j} \quad (3)$$

where \mathbf{j}_k denotes the intensity of the p_k sources in the k th parcel, δ is the Dirac distribution that allows to “shut down” the activity when the parcel is inactive ($S_k = 0$), and $\mathcal{N}(\mu_k, \Sigma_k)(\mathbf{j}_k)$ is a Gaussian distribution describing the activity within the k th parcel when it is assumed to be active ($S_k = 1$). μ_k and Σ_k represent the mean vector and the covariance matrix of the p_k sources of the k th parcel. In this study, we considered two ways of initializing the parameters of $d\mu$:

- MEM1: among each parcel, we defined: $\alpha_k = \text{Median}_{i \in i_k}(\text{APM}(i))$ and $\mu_k = 0$. i_k being the set of all the p_k sources of the k th parcel.
- MEM2: among each parcel, we defined: $\alpha_k = \text{Median}_{i \in i_k}(\text{APM}(i))$, and μ_k was initialized by the mean activity provided by the MN solution, within this parcel ($\mu_k = (1/p_k) \sum_{i \in i_k} \hat{\mathbf{J}}_{\text{MN}}(i)$).

The covariance Σ_k was set empirically, with a value on the diagonal equaling 5% of the absolute value of the mean activity

estimated using the MN solution within the parcel. Non-diagonal terms were set to zero. Under those assumptions, the MEM solution within the k th parcel $\mathbf{J}_{\text{MEM}}^k$ is found to be:

$$\mathbf{J}_{\text{MEM}}^k = \hat{\alpha}_k [\mu_k + \Sigma_k \mathbf{G}_k^T \tilde{\lambda}] \text{ with} \quad (4)$$

$$\hat{\alpha}_k = \frac{\alpha_k}{\alpha_k + (1 - \alpha_k) \exp(-F_{\mu,k}(\mathbf{G}_k^T \tilde{\lambda}))}$$

where \mathbf{G}_k and $F_{\mu,k}$ are respectively the $n \times p_k$ submatrix of \mathbf{G} and the “free energy” corresponding to the k th parcel:

$$F_{\mu,k}(\mathbf{G}_k^T \tilde{\lambda}) = \mu_k^T \mathbf{G}_k^T \tilde{\lambda} + \frac{1}{2} \tilde{\lambda}^T \mathbf{G}_k \Sigma_k \mathbf{G}_k^T \tilde{\lambda}. \quad (5)$$

Optimal $\tilde{\lambda}$ solves Eq. (B.4) where $F_{\mu}(\mathbf{G}^T \tilde{\lambda}) = \sum_{k=1}^K \log [(1 - \alpha_k) + \alpha_k \exp(F_{\mu,k}(\mathbf{G}_k^T \tilde{\lambda}))]$.

Validation

This study aims at studying the behavior of those six distributed source localization methods in a specific clinical context: the localization of EEG interictal spikes generators. To perform this validation thanks to a gold standard, we propose a fully controlled environment to generate realistic simulations of EEG data. Particular attention was paid to the choice of the anatomical location of the spike generators and their spatial extents in order to mimic realistic source configurations. Our validation objective consisted in assessing whether those six methods were able to detect those realistic spike generators with a good detection accuracy. In this section, we describe the validation data sets and validation metrics used for this evaluation study.

Validation data sets: realistic EEG simulations

Real acquisition set-up

The geometry of our simulation environment was derived from a real EEG acquisition set-up of 43 electrodes (10/10 system) performed on a patient with focal epilepsy. A high-resolution T1-weighted MRI (170 sagittal slices of 256×256 matrix, isotropic voxel size = 1 mm, $\alpha = 30^\circ$, TR = 25 ms, TE = 11 ms) was used to model the anatomy. To locate the electrodes on the anatomy, 19 among the 43 electrodes were actually acquired within the scanner and were visible on a 3D rendering of the head surface of the MRI (Gotman et al., 2004). The remaining electrodes were placed manually. The EEG acquired outside the scanner was only used to add realistic noise to the simulations. Hence, EEG background without spikes was recorded. Sampling rate was $F_s = 200$ Hz, and the signal was band pass filtered from 1.6 Hz to 35 Hz.

Head model

Realistic head models are generally required to solve the EEG forward problem, especially within the temporal lobe, because of the greater influence of the low electric conductivity of the skull. A realistic head model was obtained by segmenting the surface of the skin (12 mm mesh, 1181 nodes/2358 triangles), the skull (10 mm mesh, 922 nodes/1840 triangles) and the brain (7 mm mesh, 1185 nodes/2366 triangles) from the subject’s MRI scan. Surface segmentations were obtained using interactive thresholding and morphological methods with the Curry 4.5 software (Neuroscan, El Paso, TX) (Wagner et al., 1997). The distributed source model was obtained by segmenting the white matter/gray matter interface from

the MRI (Mangin, 1995), using the Brainvisa software¹. We described the source space by downsampling this mesh to a total of 3432 nodes (mean internodes distance = 7 ± 3 (SD) mm), the orientation of each source being set perpendicular to the cortical surface. The forward matrix \mathbf{G} was computed using the boundary element method proposed by Mosher et al. (1999), using Galerkin weighting method with linear basis functions and implemented in the Brainstorm software². Conductivities for the skin, the skull and the brain were respectively set to $0.33 \text{ S} \cdot \text{m}^{-1}$, $0.0042 \text{ S} \cdot \text{m}^{-1}$ and $0.33 \text{ S} \cdot \text{m}^{-1}$ (ratio of skull to brain of 1/80). EEG signals were considered with an average reference. The same forward problem solution \mathbf{G} was used to simulate EEG data and to perform source localization.

Spatial description of spike generators.

The spike generators were chosen in anatomical locations that are likely to occur, as observed using intracerebral recordings (Ebersole, 1997; Merlet and Gotman, 1999). We chose six locations of the spike generator by defining seed points within the temporal and the frontal lobes, namely: temporo-radial, temporo-tangential, temporo-mesial, temporo-posterior, frontal and fronto-mesial (see Fig. 1). Around each of those six seed points, spatially extended sources were generated by following the geometry of the white matter/gray matter interface mesh, using a region growing technique. Let us define by $\Theta^1(i)$, the first neighborhood order of the i th dipole on the mesh, we then define recursively the neighborhoods at higher orders ($j > 1$) as follows:

$$\Theta^j(i) = \{\Theta^{j-1}(i) \cup \{\Theta^1(i'); i' \in \Theta^{j-1}(i)\}\}, \forall j > 1, \quad (6)$$

$$\forall i \in \langle 1, p \rangle.$$

For each of the six anatomical locations, eight orders of spatial extents were generated $j \in \langle 0, 7 \rangle$, order 0 being a single dipole. These spatial extents lead to a large range of activated areas from 10 mm^2 to 40 cm^2 , as summarized in Table 1. Fig. 2 illustrates the eight spatial extents generated around the seed point chosen for the temporo-radial location.

Simulations mimicking pairs of synchronous sources with the same spatial extent (3rd neighborhood order) were also performed. The same current amplitude was set for both sources. Five combinations of paired spike generators were studied (minimal geodesic distance between both sources being indicated in each case), namely: temporo-radial + temporo-tangential (8.6 mm), temporo-radial + temporo-mesial (13.8 mm), temporo-radial + temporo-posterior (8.5 mm), temporo-radial + occipital (85.2 mm) and fronto-mesial + frontal (25.5 mm).

Simulation of the EEG interictal spike. For each of the configurations of the spike generator(s) described above, we defined a reference current distribution \mathbf{J}_{ref} that consisted of our absolute gold standard for the evaluation. The amplitude of the current was set to be uniform within the patch, showing a moment of 5 nA.m (Alarcon et al., 1994). EEG scalp potentials were then simulated by applying the forward matrix \mathbf{G} to the reference current distribution \mathbf{J}_{ref} : $\mathbf{M} = \mathbf{G} \cdot \mathbf{J}_{\text{ref}}$. Realistic noise was then added to these simulated signals. The background noise was obtained by averaging 20 independent sections of 50 ms of real EEG, randomly selected from 64 sections of real EEG background showing no spikes. Whereas all the localization methods were evaluated at one time sample, we simulated the time course of a

¹ Brainvisa: <http://www.brainvisa.info>.

² Brainstorm: <http://neuroimage.usc.edu/brainstorm/>.

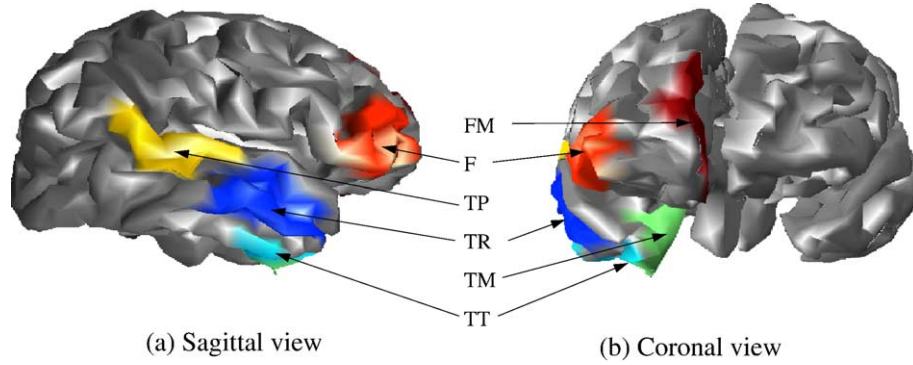


Fig. 1. Simulated sources for several anatomical locations (3rd neighborhood order, surfaces between 7 and 9 cm²): temporo-tangential (TT), temporo-mesial (TM), temporo-radial (TR), temporo-posterior (TP), frontal (F) and fronto-mesial (FM).

spike by a half-period of a sine wave lasting 50 ms for the MSP (i.e., 10 time samples at 200 Hz). Signal-to-noise ratio (SNR) was defined as the square root of the ratio of the energy of the simulated signal to the energy of the averaged noise over this 50 ms time window. Two levels of noise were generated, i.e., SNR = 6 and SNR = 2.

Simulations using addition of EEG background were also proposed in the context of spike localization in order to evaluate equivalent current dipole methods, using either averaged (Kobayashi et al., 2003) or non-averaged (Stephen et al., 2003) real EEG background. Stephen et al. (2003) showed that, when no averaging is used, such noise may contain non-epileptic physiological sources, such as α rhythm localized in the occipital lobe. In our study, the averaging of 20 sections of independent EEG background will remove any possible synchronization, providing EEG realistic noise without any physiological sources.

Validation metrics

In this section, we describe the validation metrics used to evaluate detection accuracy of the six source localization methods presented in Localization methods to be evaluated using simulated data described above. Let us denote as $\hat{\mathbf{J}}$ the solution of the inverse problem estimated by one of the six methods in one simulation context described by the reference current distribution \mathbf{J}_{ref} . The estimation of $\hat{\mathbf{J}}$ and the evaluation were performed at one single time sample.

Standard validation metrics. To assess the ability to recover the theoretical distribution of current \mathbf{J}_{ref} with an accurate amplitude, a mean square error was first estimated:

$$\text{MSE} = \frac{\sum_{i=1}^p (\hat{\mathbf{J}}(i) - \mathbf{J}_{\text{ref}}(i))^2}{\sum_{i=1}^p \mathbf{J}_{\text{ref}}(i)^2} \quad (7)$$

Im et al. (2003) suggested that the degree of focalization of a localization method may be assessed by the ratio between the estimated and theoretical energies of the current within the simulated source. Similarly, we assessed the ability of focalization by estimating an MSE only on the simulated spike generator:

$$\text{DF} = \frac{\sum_{i \in \Theta_a} (\hat{\mathbf{J}}(i) - \mathbf{J}_{\text{ref}}(i))^2}{\sum_{i \in \Theta_a} \mathbf{J}_{\text{ref}}(i)^2} \quad (8)$$

where Θ_a denotes the set of the sources belonging to the simulated spike generator.

Moreover, in order to check whether the global maximum of energy $\max(\hat{\mathbf{J}}^2)$ was accurately localized, we estimated the geodesic distance D_{min} (Im et al., 2003) between this point and the nearest point of the simulated spike generator Θ_a .

Estimation of detection accuracy. We also propose an original criterion, based on receiver operating characteristic (ROC) curves (Metz, 1986), to specifically assess detection accuracy. The conditions of application of ROC analysis are fulfilled: our simulation environment provides a gold standard \mathbf{J}_{ref} , and our detection methods involve the choice of a decision threshold, needed to build ROC curves. Detection will be defined in terms of normalized energy for both the estimated and reference current distributions for the i th source:

$$\hat{\mathbf{E}}(i) = \frac{\|\hat{\mathbf{J}}(i)\|^2 / \max(\|\hat{\mathbf{J}}\|^2)}{\max(\|\mathbf{J}_{\text{ref}}\|^2)} \quad \text{and} \quad \mathbf{E}_{\text{ref}}(i) = \frac{\|\mathbf{J}_{\text{ref}}(i)\|^2}{\max(\|\mathbf{J}_{\text{ref}}\|^2)} \quad (9)$$

For a threshold β chosen in the interval $[0, 1]$, we considered the i th source as active if its energy $\hat{\mathbf{E}}(i) \geq \beta$. By comparing the estimated activity state of each source with the one given by the gold standard \mathbf{E}_{ref} , we were able to quantify the amount of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) for each threshold $\beta \in [0, 1]$ (see Table 2). Sensitivity and specificity were then estimated as follows:

$$\text{sensitivity}(\beta) = \frac{\text{TP}(\beta)}{\text{TP}(\beta) + \text{FN}(\beta)} \quad \text{and} \quad \text{specificity}(\beta) = \frac{\text{TN}(\beta)}{\text{TN}(\beta) + \text{FP}(\beta)} \quad (10)$$

ROC curves were then obtained by plotting $\text{sensitivity}(\beta)$ against $(1 - \text{specificity}(\beta))$ for different statistical thresholds β . Detection accuracy was then assessed by the area under the ROC curve (AUC), estimated empirically. If we consider a set of randomly selected pairs of active and inactive sources, one can show that the AUC index corresponds to $\text{Prob}(\hat{\mathbf{E}}(i') > \hat{\mathbf{E}}(i))$, where i' denotes the index of an active source and i the index of an inactive source. In practice, a detection method showing an AUC index greater than 0.8 is generally considered as sufficiently accurate, allowing 80% of good detections.

Unbiased estimation of the false positive rate

For each threshold $\beta \in [0, 1]$, the number of estimated or simulated inactive sources was much greater than the number of

Table 1
Properties of the different EEG simulation contexts

Anatomical location	Neighborhood order	Surface (cm ²)	Minimum potential in μV	Electrode (min.)	Maximum potential in μV	Electrode (max.)
Temporo-radial	0	0.15	-0.5	F7	0.2	O2
	1	1.10	-2.5	F7	0.9	O2
	2	4.86	-6.0	T7	2.0	CP6
	3	8.86	-11.4	T7	3.6	C4
	4	13.92	-19.4	T7	6.4	FC6
	5	20.91	-26.3	T7	9.3	CP6
	6	28.37	-28.3	T7	11.5	CP6
Temporo-tangential	0	0.12	-0.2	T9	0.1	Fz
	1	0.67	-0.9	P9	0.7	F3
	2	2.89	-2.5	P9	2.3	F3
	3	6.96	-4.6	P9	4.4	F3
	4	13.54	-5.6	P9	5.2	F3
	5	22.21	-7.1	FT9	6.5	FC1
	6	30.95	-10.1	FT9	8.4	C3
Temporo-mesial	0	0.29	-0.2	T8	0.2	T7
	1	1.06	-1.1	F8	1.5	T7
	2	4.65	-2.2	FC6	3.4	T7
	3	8.81	-2.8	FC6	4.9	T7
	4	13.22	-3.5	T8	5.3	T7
	5	20.34	-6.2	T10	8.4	FC5
	6	31.87	-8.3	T10	10.6	CP5
Temporo-posterior	0	0.12	-0.6	CP5	0.2	T10
	1	1.26	-2.6	CP5	1.0	P10
	2	3.28	-5.1	CP5	1.7	P10
	3	7.15	-8.0	CP5	2.7	P8
	4	12.79	-15.2	CP5	5.2	FC6
	5	18.70	-23.5	CP5	8.0	FC6
	6	28.05	-32.3	CP5	11.4	T8
Frontal	0	0.16	-0.3	FC5	0.2	F8
	1	1.47	-3.3	AF3	0.9	T8
	2	3.95	-7.8	AF3	2.4	CP6
	3	7.75	-10.8	AF3	3.7	FC6
	4	15.02	-12.9	AF3	4.8	FC6
	5	22.90	-14.9	AF3	6.9	FC6
	6	30.98	-19.5	AF3	9.1	FC6
Fronto-mesial	0	0.31	-0.3	AF4	0.4	AF3
	1	0.94	-1.0	F4	1.8	AF3
	2	3.06	-4.4	AF4	3.9	AF3
	3	7.32	-9.9	AF4	5.8	FC5
	4	12.96	-14.6	AF4	8.6	FC5
	5	20.51	-17.4	AF4	10.6	FC5
	6	28.52	-18.5	AF4	11.7	FC5
7	41.08	-21.9	AF4	14.4	FC5	

For each simulated source, the name of the anatomical location and its spatial extent (neighborhood order and surface in cm²) are presented, as well as the name in the 10/10 system and the amplitude in μV of the electrodes showing the minimum and the maximum electric potentials.

active sources. Among the $p = 3432$ sources of the cortex model, generally, not more than a few hundred sources were active. To interpret the area under the ROC curve as a detection accuracy index, one should provide the same number of active and inactive sources to the ROC analysis.

Let us define by Θ the set of all the sources of the cortical mesh ($\text{card}(\Theta) = p$) and by Θ_a the set of simulated active sources ($\text{card}(\Theta_a) = p_a$ active sources). To provide less biased estimation of ROC parameters, we randomly chose a set Θ_f of p_a fictive sources among the $p - p_a$ remaining inactive sources, i.e., $\Theta_f \in \Theta \setminus \Theta_a$. Less

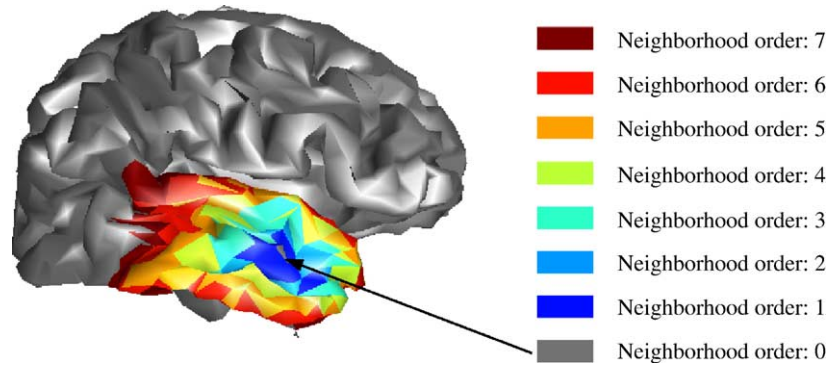


Fig. 2. Tempororadial simulated sources for several spatial extents, from neighborhood order 0 (15 mm²) to 7 (36 cm²).

biased ROC curves and AUC were then estimated for those sets of p_a pairs of active Θ_a and fictive Θ_f sources.

However, the false positive rate may now be largely underestimated because many spurious sources are missed by the random drawing of Θ_f in $\Theta \setminus \Theta_a$. We thus proposed two strategies to choose Θ_f , optimizing the false positive detection. We decided to split the estimation of detection accuracy in two components, one dedicated to the focalization ability of the method and the second dedicated to false positive detection far from the simulated spike generators:

- Focalization ability: a criterion AUC_{close} was estimated by choosing the fictive sources within a neighborhood of the simulated spike generator, i.e., $\Theta_f \in \Theta^{10}(\Theta_a) \setminus \Theta_a$, where $\Theta^{10}(\Theta_a)$ denotes the 10th neighborhood order of Θ_a , as defined in Eq. (6). AUC_{close} quantifies the detection ability in a close neighborhood of the simulated spike generator, assessing whether the activity and its spatial extent were accurately recovered or not. Any over-estimation or under-estimation of the spatial extent of the source should then affect this criterion.
- Detection of false positive far from the spike generator: a criterion AUC_{far} was estimated by choosing the fictive sources among local maxima of activity $\hat{\mathbf{E}}$ located far from to the simulated spike generator, i.e., within the complementary set of $\Theta^{10}(\Theta_a)$: $\Theta_f \in \Theta \setminus \Theta^{10}(\Theta_a)$. To ensure that the detection of all local maxima of $\hat{\mathbf{E}}$ are uniformly distributed over the cortex, we used the cortex parcelling defined in the MEM approach (see Localization methods to be evaluated). Among each of the K parcels located far from the spike generator ($\Theta \setminus \Theta^{10}(\Theta_a)$), the sources were sorted according to their decreasing energy $\hat{\mathbf{E}}(i)$. To set Θ_f , we randomly select a parcel among K and then select the most active source within that parcel and so on until we select p_a fictive sources. Occurrence of false positive generated far from the simulated spike generator should then reduce the criterion AUC_{far} , even if some activity was accurately detected within the simulated spike generator.

The proposed index of detection accuracy AUC was then defined as the mean of the previous criteria:

$$AUC = \frac{1}{2} (AUC_{\text{close}} + AUC_{\text{far}}) \quad (11)$$

To obtain consistent measurements not sensitive to one particular choice of Θ_f , AUC was estimated over 50 independent drawings of Θ_f , and the mean AUC over those 50 trials will be presented here as an index of detection accuracy.

Effect of source localization spatial resolution

When very focal spike generators were simulated (cf. 0 or 1 neighborhood orders), our estimation of AUC may still be biased because too few pairs of active and fictive sources were considered for the estimation. Moreover, the ROC analysis considered that the detection failed if the estimated activity was not found exactly on the very small area of the simulated generator, which may be over conservative. We thus decided to take into account the spatial resolution of the localization methods when evaluating detection accuracy.

For very focal simulated generators, we slightly increase the set of active sources. In practice, if a focal activity is found in the close neighborhood of such a focal generator, let us say around 1 cm away, this detection should still be considered a success. When the spike generator consisted of a single dipole (neighborhood order: 0), Θ_a was replaced by its 2nd neighborhood order $\Theta^2(\Theta_a)$, that is, all the sources of $\Theta^2(\Theta_a)$ were considered as theoretically active for the ROC analysis. When the spike generator consisted of a 1st neighborhood order, Θ_a was replaced by its 1st neighborhood order $\Theta^1(\Theta_a)$. Using such a convention, the smallest set of activated sources corresponded to an average area of 3–4 cm² (see Table 1), i.e., $p_a \approx 24$ pairs of active and fictive sources. Such an area was considered as a low bound of the spatial resolution of the different source localization methods evaluated here.

Table 2

Definition of the parameters needed to generate the ROC curves

	Reference inactive sources $\mathbf{E}_{\text{ref}}(i) = 0$	Reference active sources $\mathbf{E}_{\text{ref}}(i) = 1$
Estimated inactive sources $\hat{\mathbf{E}}(i) < \beta$	True negative TN (β)	False negative FN (β)
Estimated active sources $\hat{\mathbf{E}}(i) \geq \beta$	False positive FP (β)	True positive TP (β)

Each case of the table refers to the number of observed sources matching both criteria, for a given detection threshold $\beta \in [0, 1]$.

Results

In this section, we present the evaluation of the six source localization methods using standard validation metrics (MSE, DF and Dmin) and our detection accuracy index AUC. We also include the evaluation of the APM maps using AUC and Dmin to assess the behavior of this preprocessing step. For all methods derived from the Bayesian framework, the hyperparameter α was estimated by the empirical approach of the “L curve” (Hansen, 1992;

Gorodnitsky et al., 1995). For the MEM-based methods, the variance of the noise σ_E^2 in Eq. (B.4) was estimated from the average background activity before we added any signal.

Visual evaluation of some localization results

We first checked the agreement between our detection accuracy index AUC and the visual analysis of the source localization results. To visualize the results, we show the absolute value of the reconstructed activity $|\hat{\mathbf{J}}|$ thresholded upon the level of background activity (Otsu, 1979) (see Figs. 3–5).

Complete mislocalizations were clearly detected by the criterion AUC, as illustrated in Fig. 3, where we show that LORETA1 completely missed the source (AUC = 0.38, for a temporo-radial source, 3rd neighborhood order). Such behavior was probably due to some numeric instabilities caused by the computation of a discrete Laplacian on a complex cortical surface (David and Garnero, 2002). The additional regularization using a minimum norm term implemented in LORETA2 increased the stability of the method, leading to very accurate localization result (AUC = 0.97). Moreover, most of the localization methods were able to recover correctly sources of different spatial extents, except for the mesial anatomical locations, i.e., temporo-mesial and fronto-mesial. When visualizing current amplitudes $\hat{\mathbf{J}}$ (results not shown), positive current within the temporo-mesial area was most often estimated by negative current on the temporo-radial area. Similarly, fronto-mesial positive current was most often estimated by bilateral frontal currents, one positive and the other negative.

In order to show the influence of the source spatial extent, Figs. 4 and 5 present MEM2 and LORETA2 results for temporo-radial and temporo-tangential sources, for three spatial extents (i.e., 2nd, 4th and 7th neighborhood orders). Although absolute current amplitude was generally underestimated by the MEM2 approach, both methods performed quite well in terms of detection accuracy. For almost all the presented simulations, we found a good agreement between our detection accuracy index $AUC > 0.93$ and visual interpretation of the results. Moreover, the spurious source generated by MEM2 within the sylvian fissure in Fig. 5h effectively lead to a lower AUC of 0.76. Those trends were confirmed by the following quantitative analysis.

Standard validation metric

Fig. 6 presents the distributions of the standard validation metrics (DF, MSE and Dmin) for all simulations performed with one single source for both noise levels $SNR = 6$ and $SNR = 2$. As very similar behavior was observed for all anatomical locations investigated and for all spatial extents, we pooled the results, showing just the distribution of each metric using boxplot representations. Boxplot is a standard representation to display the distribution of an index by showing its median, first and third quartiles to create the box, as well as isolated outliers using single points.

Figs. 6a and b show that LORETA-based methods estimated the current amplitudes within the focus more accurately than the others (lower DF). However, those methods were also the worst to recover accurate current amplitude over the whole brain (higher MSE), suggesting either the presence of spurious sources or a compensation of the positive current estimated within the focus by a surrounding negative current. LORETA2 is less sensitive to this effect, leading to lower MSE values, whereas higher noise level clearly increases MSE for both LORETA1 and LORETA2. Moreover, we observed that MN-based and MEM-based methods underestimated current amplitude, leading to values of MSE and DF very close to 1.

Occurrence of spurious sources was studied by measuring the minimal geodesic distance Dmin between the global maximum of reconstructed activity $\max(\hat{\mathbf{J}}^2)$ and the simulated sources, as shown in Figs. 6c and d. Geodesic distance following the circinvoluted surface of the cortex is indeed more discriminant than Euclidian distance. For instance, sources mislocalized close to the sylvian fissure or close to the interhemispheric plane may reflect a very small Euclidian distance and a very large geodesic distance. The measured highest values of Dmin actually correspond to a localization of the global maximum of activity on the wrong hemisphere for frontal and fronto-mesial simulated sources or on the wrong side of the sylvian fissure for temporo-posterior sources. In most of those cases, the Euclidian distance was less than 2 cm. Table 3 presents the number of cases over 8 spatial extents where Dmin was lower than 2 cm. Results show that LORETA-based and MEM-based methods provided spurious sources showing maximum activity located very far from the

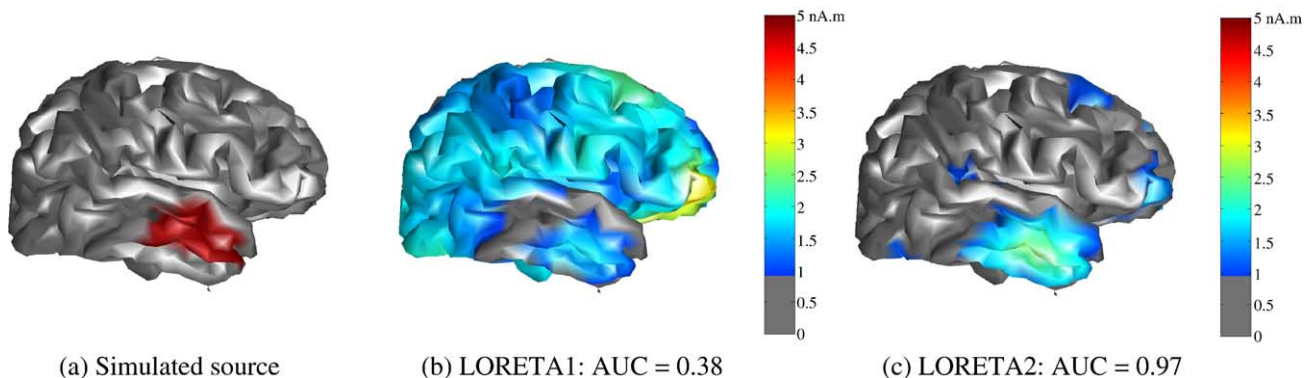


Fig. 3. Mislocalization of LORETA1 due to some numerical instabilities: (a) simulated source (3rd neighborhood order of the temporo-radial source), (b) LORETA1 and (c) LORETA2 (simulations with $SNR = 6$). For each localization, the absolute value of the reconstructed activity $|\hat{\mathbf{J}}|$ is presented using the same color coding between 0 and 5 nA.m, with the corresponding AUC index. The local maximum of activity observed in the frontal lobe for LORETA1 was actually an important negative current, “trying” to compensate a positive current within the temporal lobe. The additional regularization used for LORETA2 allowed to avoid such instability leading to very accurate localization (AUC = 0.97).

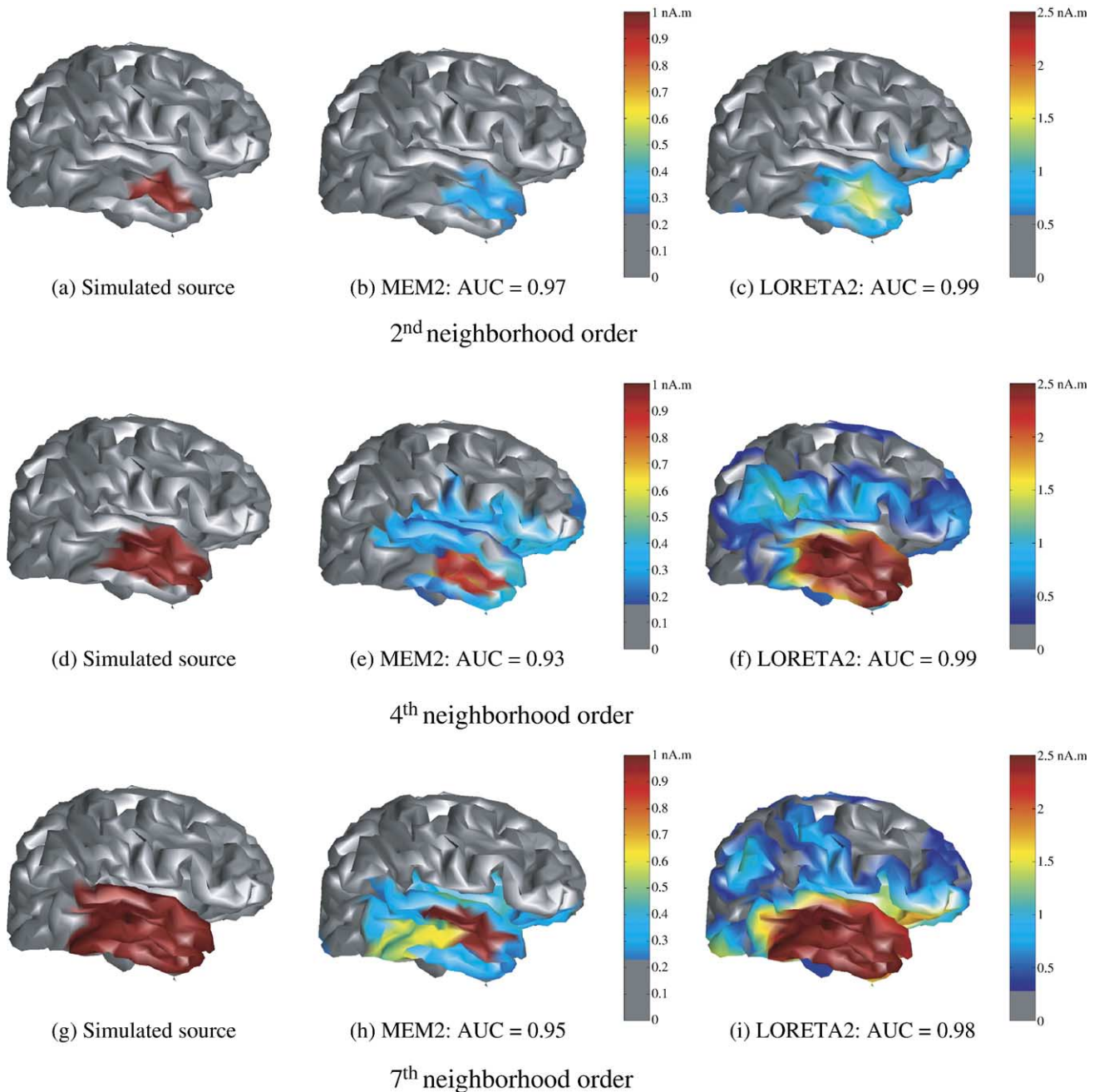


Fig. 4. Simulation of a single temporo-radial source with noise level SNR = 6. Illustration of the simulated source as well as MEM2 and LORETA2 localization results for three spatial extents (i.e., 2nd, 4th and 7th neighborhood orders). In each case, the absolute value of the reconstructed activity $|\hat{J}|$ is presented with the corresponding AUC index, using the same color coding between 0 and 2.5 nA.m for LORETA2 and between 0 and 1 nA.m for MEM2. Both methods recover simulated sources accurately, whereas LORETA2 seemed slightly more accurate. However, LORETA2 also detected some spurious sources in the parietal and occipital lobes (see f and i).

simulated source (e.g., on the contralateral hemisphere, on the occipital lobe), whereas APM and WMN were the methods which located the global maximum of activity the most accurately. Table 3 shows that APM was the only method able to locate accurately the maximum activity of the temporo-mesial and fronto-mesial sources, with SNR = 6. At higher noise level (SNR = 2), we observed an increase of D_{\min} for most methods, suggesting that more spurious sources were generated as it is expected. However, we note that MEM2 was the method the least sensitive to the level of noise.

Evaluation of detection accuracy

Considering $AUC \geq 0.8$ as a good detection accuracy, results show that very accurate detection (mean $AUC \geq 0.9$ over the 8 spatial extents) was obtained by at least one of the six localization methods for each anatomical location, except for temporo-mesial and fronto-mesial sources (Table 4). For each localization method and each noise level, the distribution of AUC over all simulated sources is presented in Fig. 7a. Analysis of variance shows a significant effect of the type of localization method over detection

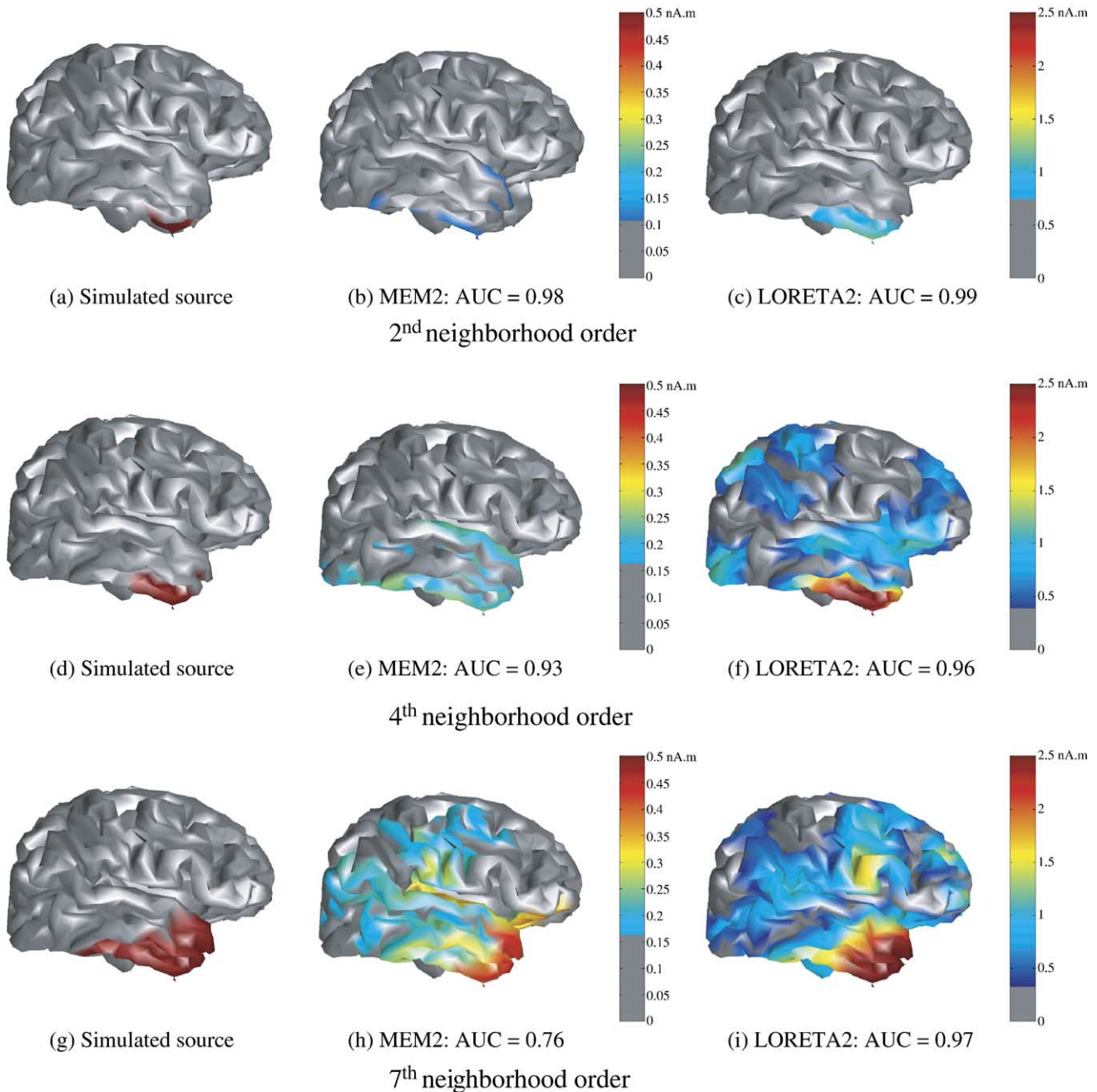


Fig. 5. Simulation of a single temporo-tangential source with noise level $\text{SNR} = 6$. Illustration of the simulated source as well as MEM2 and LORETA2 localization results for three spatial extents (i.e., 2nd, 4th and 7th neighborhood orders). In each case, the absolute value of the reconstructed activity $|\hat{\mathbf{J}}|$ is presented with the corresponding AUC index, using the same color coding between 0 and 2.5 nA.m for LORETA2 and between 0 and 0.5 nA.m for MEM2. LORETA2 seemed to be more accurate than MEM2 to recover temporo-tangential sources. However, both methods showed some spurious sources with potentially high activity. See for instance the source generated by MEM2 within the sylvian fissure in panel h leading to an AUC of 0.76.

accuracy AUC ($F = 19.2$, $P < 0.001$ and $R_{\text{adjust}}^2 = 0.26$). However, the observed low adjusted determination coefficients R_{adjust}^2 was due to some very low AUC values considered as outliers, as observed in particular for LORETA-based methods (Fig. 7a). Those outliers may be explained by numerical instabilities as observed for LORETA1 (Fig. 3) and by a higher sensitivity to the noise level resulting in more spurious sources for both LORETA1 and LORETA2. Post-hoc paired comparisons show that MEM2 for both noise levels (t test, Bonferroni corrected $P < 0.005$) and LORETA2 at $\text{SNR} = 6$ (t test, Bonferroni corrected $P < 0.05$) were significantly more accurate

than MN. APM maps were significantly less accurate than MN (t test, Bonferroni corrected $P < 0.001$). This last result is not surprising since those probability-like coefficients of activation were intended to be very sensitive in order not to miss any potentially active source, but to the price lower specificity.

Effect of the noise level on detection accuracy

For each method, we have tested the influence of the noise level on the mean detection accuracy AUC, using a non-parametric

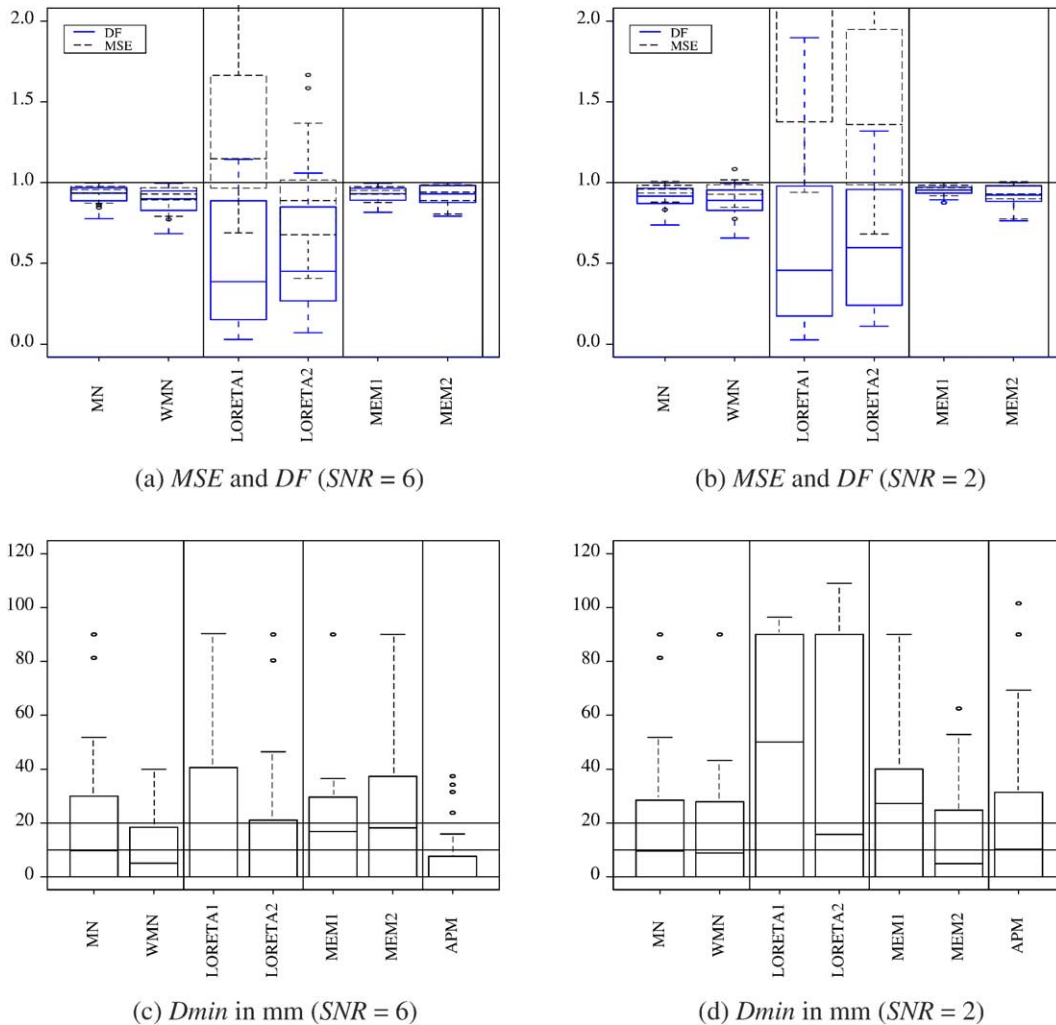


Fig. 6. Standard validation metrics for simulations involving one single source: distribution of DF (plain blue) and MSE (dashed black) for signal-to-noise ratios SNR = 6 (a) and SNR = 2 (b), distribution of Dmin in mm for signal-to-noise ratios SNR = 6 (c) and SNR = 2 (d) (horizontal lines correspond to Dmin = 10 mm and 20 mm). Distributions, shown using boxplot representations, correspond to the results for all anatomical locations and spatial extents pooled together. LORETA-based methods estimated the current amplitudes within the focus more accurately than the others (lower DF), whereas they were the worst methods to recover accurate current amplitude on the whole brain (higher MSE), and even worst when increasing the noise level. This observation suggests the occurrence of spurious sources, as confirmed by higher Dmin values. Note that both APM and WMN localized the maximum of activity quite accurately most of the time (lower Dmin), whereas increasing the noise level clearly lowered those performances.

paired t test with bootstrap resampling (Efron and Tibshirani, 1993). Reducing the signal-to-noise ratio from SNR = 6 to SNR = 2 leads to a highly significant decrease of the mean AUC for both WMN and MEM1 (Bonferroni corrected $P < 0.001$) and to a slightly significant decrease for LORETA2 (Bonferroni corrected $P < 0.01$). Comparing the distribution of AUC at both noise levels in Fig. 7a showed that MEM2 was again the method the least sensitive to the level of noise.

Effect of the anatomical location and spatial extent on detection accuracy

When pooling all the localization methods and spatial extents, we found a clear effect of anatomical location over the mean AUC ($F = 11.77$, $P < 0.001$ and $R^2_{\text{adjust}} = 0.15$), as shown in Fig. 7b. Post-hoc paired comparisons show that source localization involving the temporo-mesial structures for both noise levels (t test, Bonferroni

corrected $P < 0.005$), and the fronto-mesial structures at SNR = 6 (t test, Bonferroni corrected $P < 0.05$) lead to significantly less accurate results than for the other anatomical locations. When pooling all the localization methods and anatomical locations, distributions of AUC for each spatial extent were very widespread (see Fig. 7c), and no effect could be interpretable by an analysis of variance ($R^2_{\text{adjust}} = 0.026$). Even if AUC seemed slightly lower for the largest simulated sources (see 7th neighborhood order in Fig. 7c), no clear trend of an effect of the spatial extent over AUC was observed when analyzing those results for each localization method and for each anatomical location (results not shown).

Fig. 8 and Table 4 provide a more detailed analysis of the effect of anatomical location. The distribution of AUC over the 8 spatial extents is presented for each localization method, for each noise level and for each anatomical location, using boxplot representations. Results confirm the previous trends observed on the global analysis. Comparing LORETA1 and LORETA2 for the temporo-

Table 3

Number of good localizations of the global maximum of activity, i.e., when $D_{min} < 2$ cm, over the 8 spatial extents for each anatomical location and each signal-to-noise ratio (SNR = 6 top and SNR = 2 bottom)

Anatomical location	APM	MN	WMN	LORETA1	LORETA2	MEM1	MEM2
<i>SNR = 6: # $D_{min} < 2$ cm</i>							
Frontal	6	7	6	7	8	7	5
Fronto-mesial	8	4	8	2	5	2	1
Temporo-mesial	8	2	2	2	2	2	0
Temporo-posterior	8	4	5	7	7	4	6
Temporo-radial	6	7	8	4	6	8	8
Temporo-tangential	8	3	8	8	8	3	5
<i>SNR = 2: # $D_{min} < 2$ cm</i>							
Frontal	5	8	7	5	6	7	8
Fronto-mesial	6	5	7	0	3	0	7
Temporo-mesial	7	0	0	1	0	1	1
Temporo-posterior	4	4	5	5	6	4	5
Temporo-radial	4	7	7	2	4	7	8
Temporo-tangential	6	4	6	7	7	3	5

Most accurate methods are displayed using bold font. Note that both APM and WMN localized the maximum of activity very accurately in almost all configurations. We observed less good localizations of the maximum of reconstructed activity at higher noise level (SNR = 2), suggesting that more spurious sources were generated.

radial sources (Fig. 8a) clearly shows that the additional MN regularization implemented in LORETA2 allowed to avoid many numerical instabilities of LORETA1. However, all LORETA-based methods were clearly sensitive to the level of noise. The main result was that good detection accuracy (AUC > 0.8) was observed

at least by LORETA2 and MEM2 for all anatomical locations except for mesial structures (Figs. 8c and f). Increasing the noise level clearly decreased detection accuracy for all methods, except for MEM2.

Simulation of pairs of synchronous sources

Table 5 presents the results of the evaluation when pairs of synchronous sources were simulated (SNR = 6, 3rd neighborhood order spatial extent). For each simulation, we present the global detection accuracy index AUC, and for each simulated source of the pair, the local detection accuracy index AUC_{close} and the geodesic distance D_{min} to the closest local maximum of $|\hat{\mathbf{J}}|$. For each index AUC_{close} , the dipoles belonging to the other simulated source were included neither in Θ_a nor in Θ_f (cf. Unbiased estimation of the false positive rate). A dipole i was considered as a local maximum if $|\hat{\mathbf{J}}(i)|$ was greater than its neighbors and if $|\hat{\mathbf{J}}(i)| > 0.5 \times \max(|\hat{\mathbf{J}}|)$, that is, local maxima were selected among the full width at half maximum of the distribution of $|\hat{\mathbf{J}}|$. Our global detection accuracy index AUC was most of the time in agreement with local indexes AUC_{close} and D_{min} assessed for each simulated source. We observed similar trends to those obtained with simulations involving one single source, notably, the different localization methods provided very complementary qualities. LORETA-based and MEM-based approaches were among the most accurate in terms of AUC and AUC_{close} , whereas only APM was able to find a local maximum within or close to each simulated source (cf. $D_{min} < 7.7$ mm). AUC_{close} and D_{min} results show that temporo-mesial and fronto-mesial sources were hardly recovered by most of the localization methods. Two failures of LORETA1 due to numerical instabilities were observed (temporo-mesial + temporo-radial simulation: AUC = 0.667 and temporo-tangential + temporo-radial simulation: AUC = 0.298) and confirmed by higher D_{min} . An additional source within the occipital lobe was added to study the ability of the methods to detect very distant and orthogonal sources (temporo-radial + occipital). Both sources were clearly identified by LORETA1, LORETA2 and MEM2, whereas LORETA2 slightly mislocalized the temporo-radial source leading

Table 4

Estimation of AUC for each simulation context and each signal-to-noise ratios (SNR = 6 top and SNR = 2 bottom)

Anatomical location	APM	MN	WMN	LORETA1	LORETA2	MEM1	MEM2
<i>SNR = 6: AUC</i>							
Frontal	0.505 (0.1466)	0.802 (0.0791)	0.769 (0.1070)	0.982 (0.01126)	0.985 (0.0141)	0.812 (0.1017)	0.888 (0.0908)
Fronto-mesial	0.488 (0.1659)	0.756 (0.1053)	0.715 (0.1270)	0.667 (0.28717)	0.731 (0.2518)	0.747 (0.1101)	0.887 (0.0888)
Temporo-mesial	0.562 (0.1469)	0.656 (0.0508)	0.671 (0.0843)	0.554 (0.27627)	0.512 (0.2735)	0.638 (0.0714)	0.733 (0.0938)
Temporo-posterior	0.461 (0.0854)	0.725 (0.0888)	0.690 (0.1111)	0.863 (0.25983)	0.943 (0.0753)	0.751 (0.0952)	0.935 (0.0325)
Temporo-radial	0.552 (0.0589)	0.803 (0.0297)	0.768 (0.0514)	0.692 (0.39887)	0.943 (0.0876)	0.788 (0.0359)	0.916 (0.1062)
Temporo-tangential	0.578 (0.0907)	0.713 (0.0351)	0.765 (0.0438)	0.989 (0.00804)	0.982 (0.0140)	0.736 (0.0608)	0.875 (0.0903)
<i>SNR = 2: AUC</i>							
Frontal	0.532 (0.1505)	0.801 (0.0775)	0.735 (0.0783)	0.870 (0.1303)	0.905 (0.1003)	0.745 (0.1171)	0.856 (0.1186)
Fronto-mesial	0.486 (0.1127)	0.717 (0.0894)	0.654 (0.1197)	0.525 (0.2145)	0.622 (0.2254)	0.690 (0.1298)	0.882 (0.1017)
Temporo-mesial	0.603 (0.1145)	0.572 (0.0838)	0.563 (0.0636)	0.441 (0.1906)	0.446 (0.2355)	0.538 (0.1139)	0.707 (0.1420)
Temporo-posterior	0.507 (0.0718)	0.716 (0.0703)	0.658 (0.0733)	0.852 (0.0946)	0.882 (0.0787)	0.689 (0.1076)	0.917 (0.0314)
Temporo-radial	0.570 (0.0675)	0.774 (0.0581)	0.719 (0.0673)	0.599 (0.2929)	0.748 (0.2482)	0.750 (0.0774)	0.941 (0.0427)
Temporo-tangential	0.577 (0.1967)	0.696 (0.0352)	0.705 (0.0927)	0.966 (0.0344)	0.964 (0.0352)	0.645 (0.0624)	0.918 (0.0693)

For each anatomical location, mean (standard deviation) of AUC values over the 8 spatial extents is shown. Most accurate methods, showing the highest mean AUC, are displayed using bold font. Except for the mesial anatomical locations, LORETA2 and MEM2 were the most accurate methods, whereas LORETA1 was the least robust (cf. higher standard deviations). Note that on average all AUC decreased when SNR decreased from 6 to 2, whereas MEM2 seemed to be the most robust method regarding the level of noise.

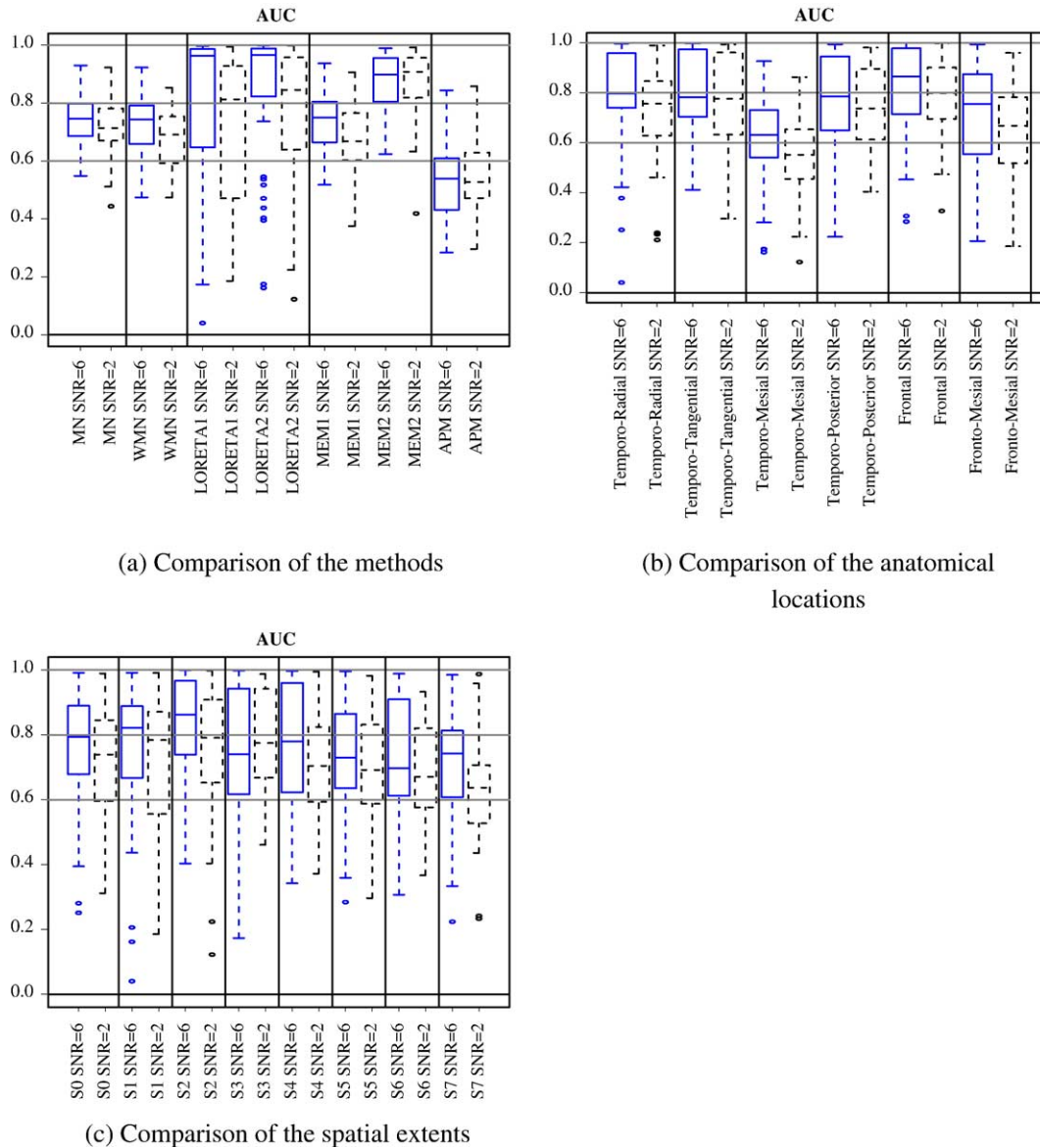


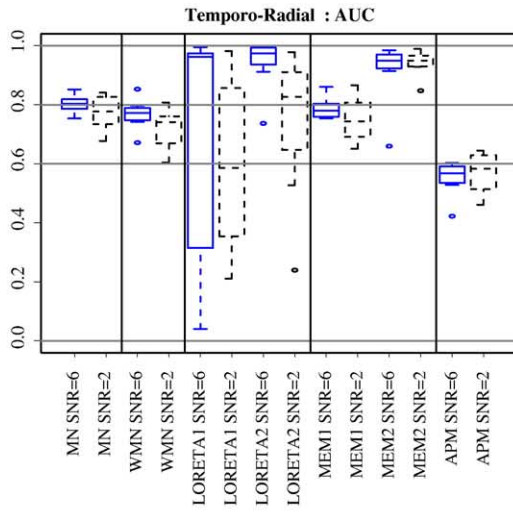
Fig. 7. ROC analysis for simulations of a single source. Distribution of AUC for both signal-to-noise ratios (SNR = 6 in plain blue and SNR = 2 in dashed black): (a) distribution of AUC over the 8 spatial extents and all anatomical locations for each localization method, (b) distribution of AUC over the 8 spatial extents and all localization methods for each anatomical location and (c) distribution of AUC over all anatomical locations and all localization methods for each spatial extent, from neighborhood order 0 (S0) to 7 (S7) (horizontal lines correspond to AUC = 0.6; 0.8 and 1.0). We observed a significant effect of the type of localization method and of the anatomical location of the sources on detection accuracy AUC, whereas no effect of the spatial extent was found. Most accurate methods were LORETA2 and MEM2, whereas LORETA1 was the least robust method. All methods were significantly less accurate when recovering temporo-mesial and fronto-mesial sources. Increasing the noise level clearly decreased detection accuracy in almost all situations except for MEM2.

to a lower detection accuracy (AUC = 0.75). By measuring D_{min} , we also noticed that LORETA-like methods better identified the occipital source, whereas all the other methods better recovered the temporo-radial source. However, our criteria to define local maxima (cf. $|\hat{J}(i)| > 0.5 \times \max(|\hat{J}|)$) were certainly too selective, notably for LORETA-like methods, whose smooth distributions showed very few local maxima.

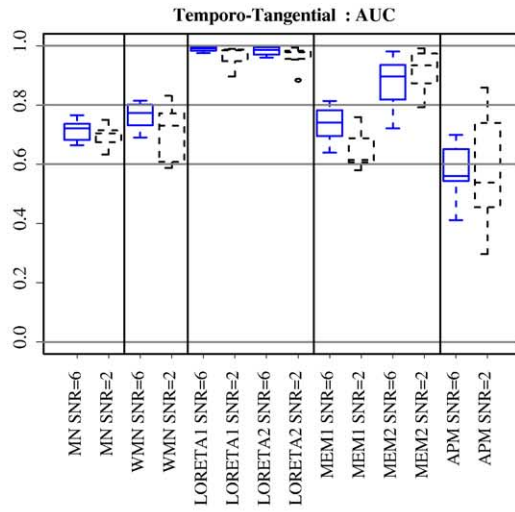
Discussion

Following the validation methodology described in Jannin et al. (2002), we have proposed an original validation framework to assess detection accuracy of distributed source localization methods in the clinical context of interictal spike localization. Such a framework consists in generating realistic validation data

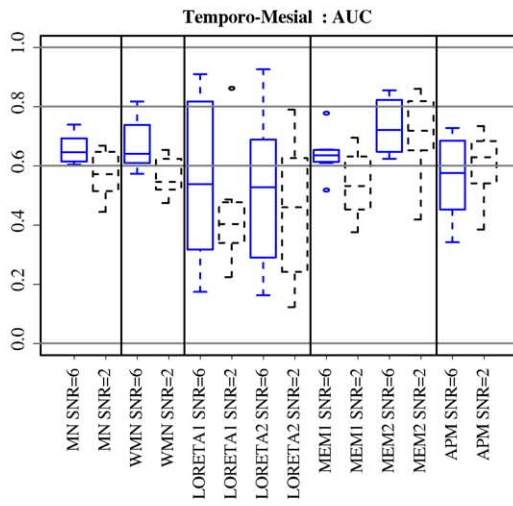
Fig. 8. ROC analysis for simulations of a single source. Distribution of AUC for both signal-to-noise ratios (SNR = 6 plain blue and SNR = 2 dashed black) over the 8 spatial extents, for each anatomical location. This detailed analysis confirmed the trends observed on the previous global ones. LORETA2 and MEM2 were the most accurate methods, whereas none of the method accurately recovered the temporo-mesial sources (horizontal lines correspond to AUC = 0.6; 0.8; and 1.0).



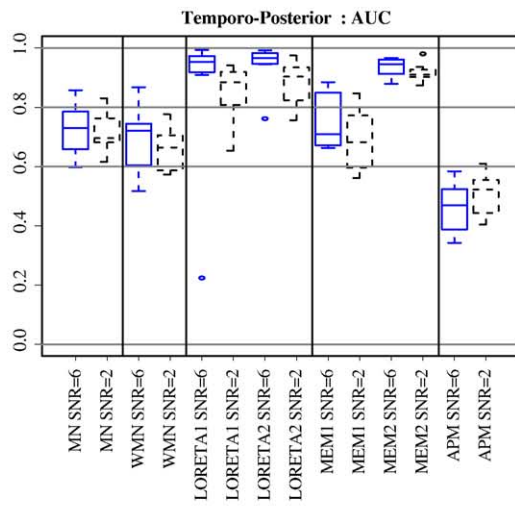
(a) Temporo-Radial



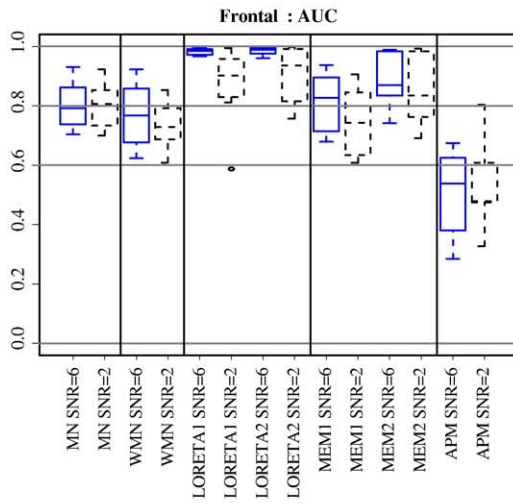
(b) Temporo-Tangential



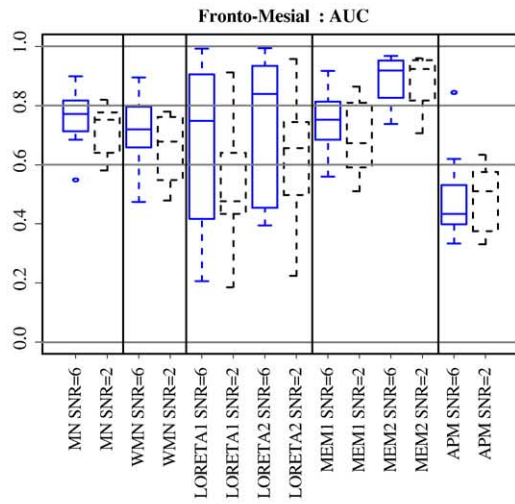
(c) Temporo-Mesial



(d) Temporo-Posterior



(e) Frontal



(f) Fronto-Mesial

Table 5
Simulation of pairs of synchronous sources (SNR = 6, spatial extent: 3rd neighborhood order)

Anatomical locations	APM		MN		WMN		LORETA1		LORETA2		MEM1		MEM2	
	AUC _{close}	Dmin	AUC _{close}	Dmin	AUC _{close}	Dmin	AUC _{close}	Dmin	AUC _{close}	Dmin	AUC _{close}	Dmin	AUC _{close}	Dmin
Fronto-mesial	0.614	0.0	0.597	9.1	0.509	62.0	0.816	–	0.717	–	0.618	62.0	0.674	–
+Frontal	0.451	0.0	0.803	0.0	0.766	0.0	0.952	0.0	0.940	0.0	0.791	0.0	0.703	75.7
AUC on all sources	0.295		0.621		0.532		0.949		0.882		0.624		0.737	
Temporo-mesial	0.494	0.0	0.626	–	0.560	–	0.408	37.6	0.755	–	0.622	–	0.797	–
+Temporo-radial	0.468	0.0	0.713	0.0	0.669	0.0	0.690	–	0.978	0.0	0.742	0.0	0.890	0.0
AUC on all sources	0.363		0.750		0.652		0.667		0.904		0.780		0.918	
Temporo-radial	0.453	0.0	0.705	0.0	0.630	0.0	0.949	–	0.941	–	0.713	0.0	0.847	0.0
+Occipital	0.385	7.7	0.664	–	0.640	–	0.970	0.0	0.977	0.0	0.673	–	0.813	–
AUC on all sources	0.246		0.652		0.597		0.972		0.750		0.670		0.878	
Temporo-radial	0.610	0.0	0.726	0.0	0.715	0.0	0.979	0.0	0.985	0.0	0.716	0.0	0.944	0.0
+Temporo-posterior	0.471	0.0	0.671	0.0	0.615	0.0	0.838	–	0.857	–	0.654	0.0	0.782	–
AUC on all sources	0.400		0.765		0.679		0.958		0.959		0.743		0.933	
Temporo-tangential	0.644	0.0	0.638	0.0	0.662	–	0.136	50.6	0.653	3.2	0.622	0.0	0.726	0.0
+Temporo-radial	0.610	0.0	0.762	0.0	0.721	0.0	0.281	–	0.889	0.0	0.764	0.0	0.879	0.0
AUC on all sources	0.598		0.754		0.732		0.298		0.813		0.760		0.882	

For each simulation, estimation of the global detection accuracy index AUC. For each simulated source of the pair, estimation of the local detection accuracy index AUC_{close} and of the geodesic distance Dmin in millimeters to the closest local maximum of $|\hat{J}|$. ‘–’ indicates that no local maxima greater than $0.5 \times \max(|\hat{J}|)$ was found in the neighborhood of this source. Most accurate methods regarding AUC are displayed using bold font. Compared to single source simulations, main trends were confirmed: LORETA1 showed some numerical instabilities for two simulations, MEM2 and LORETA2 were the most accurate methods and APM localized local optima more accurately. Combinations temporo-tangential + temporo-radial and temporo-mesial + temporo-radial were the most difficult situations to achieve good detection accuracy.

sets, provided by interictal spike simulations, mimicking realistic source configurations involving several anatomical locations and spatial extents. The main methodological contribution of this paper consisted in the criterion AUC we proposed to assess detection accuracy. We adapted the well-known area under the ROC curve criterion (Metz, 1986) to the context of distributed source localization methods, focusing on a more accurate detection of the false positive rate. This new validation metric, associated with more standard ones (cf. MSE, DF and Dmin), was used to evaluate and compare six source localization methods.

Realistic validation data sets

Validation data sets providing an absolute gold standard for validation purposes were generated using realistic simulations of EEG interictal spikes. A real EEG acquisition set-up associated with an anatomical MRI was used to define the position of the electrodes, the distributed source model and the head model. Realistic source configurations, mimicking activity patterns observed with intracerebral recordings, were then generated. The source configurations involved six regions of the temporal and frontal lobes. These regions were quite similar to those simulated by Kobayashi et al. (2003). For each anatomical location, our main contribution consisted in simulating a wide range of spatial extents, from 10 mm² to 40 cm². Producing realistic EEG simulations of interictal spikes has mainly been performed to study equivalent current dipole-based methods (Stephen et al., 2003; Kobayashi et al., 2003; Bénar et al., 2005) that can at best estimate the center of mass of an extended source. The purpose of this study was to evaluate the ability of distributed source localization methods to recover spatially extended sources. Whereas 6 cm² of active cortex is usually considered as a low bound to generate a spike on the

scalp (Ebersole, 1997), most evaluation studies using extended sources actually simulated one or several cortical patches of areas ranging from 1 cm² to 6 cm² (Schmidt et al., 1999; David and Garnero, 2002; Stephen et al., 2003; Kobayashi et al., 2003; Im et al., 2003; Yamashita et al., 2004; Trujillo-Barreto et al., 2004; Bénar et al., 2005). None of those studies actually addressed the problem of detecting very widespread sources (around 30 or 40 cm²), whereas this may occur for real spikes (Ebersole, 1997).

Modeling the physical aspects of spike generation on scalp EEG was then performed by solving the EEG forward problem using the boundary element method (Mosher et al., 1999). Simulated EEG signals were then corrupted by noise by adding average real EEG background showing no epileptic activity as proposed in Kobayashi et al. (2003) and Bénar et al. (2005). Random sampling and averaging of real EEG background segments removed any possible synchronization, allowing to produce spatially correlated noise without any physiological source.

Evaluation of EEG/MEG localization methods using realistic simulations is a useful framework as it allows to study the impact of many parameters on the system. We focused on realistic source configurations for interictal spikes, using our usual clinical setting (cf. 43 EEG electrodes). Studying the effect of the number of electrodes, the modality (EEG and/or MEG), the spatial resolution of the source model (i.e., number of sources p) and the conductivities of the head model on the localization of the sources is feasible within such a framework, but it was out of the scope of this study. Some of those aspects were studied in detail for single dipoles (Fuchs et al. (1998), Pascual-Marqui (2002)) or small extended sources (Liu et al., 2002; Trujillo-Barreto et al., 2004) simulations. The simulation set-up proposed here and notably the head and source models may be slightly inaccurate due to

segmentation errors. For instance, using $p = 3432$ nodes for the cortical surface (mean internodes distance = 7 ± 3 (SD) mm) may be too coarse to estimate accurately some cortical normals and should be considered as a minimum spatial sampling in practice. However, in this evaluation study, the same bias influenced all the methods in the same way, therefore it is still valid to compare those methods.

Validation metrics for detection accuracy

Whereas evaluation metrics for equivalent current dipoles methods are well-established (Baillet et al., 2001), quantifying the performance of distributed source methods is far from trivial, and no consensus is actually accepted (Grave de Peralta Menendez and Gonzales Andino, 1998; Liu et al., 2002; Im et al., 2003). In addition to the well known metric MSE, we used two validation metrics, DF and Dmin, similar to those recently proposed by Im et al. (2003). An important contribution of the present work was to propose an original index of detection accuracy AUC that proved complementary and more discriminant compared to the other metrics (MSE, DF and Dmin).

Comparing the amplitudes between $\hat{\mathbf{J}}$ and \mathbf{J}_{ref} , MSE provides a global measurement that may be confusing, notably when assessing solutions relatively far from the simulated current distribution. On the other hand, DF assesses specifically the amount of energy recovered within the simulated source, as a measure of focalization ability. However, energy of the reconstructed current distribution $\hat{\mathbf{J}}$ is not the optimal way to assess detection ability. Studying the distribution and notably the location of the peaks of activity (local maxima) is more reliable. Therefore, we first use Dmin to measure minimal geodesic distance between the global maximum of the reconstructed activity and the simulated source (Im et al., 2003). Following the circonvoluted structure of the cortical surface, geodesic distance is able to detect many mislocalizations (as sources located on the wrong side of a sulcus, on the wrong side of the sylvian fissure or on the wrong hemisphere), whereas the Euclidian distance may be small in such situations. This issue is particularly important when using distributed source models with fixed dipole orientations. Even if we were able to detect spurious sources using Dmin, we only took into account local maxima of reconstructed activity when analyzing simulations with pairs of synchronous sources. Only the global maximum of reconstructed activity was used for single source simulations. However, all local maxima of reconstructed activity and corresponding spatial extents should be considered to provide an appropriate measure of detection accuracy.

The ability of distributed source localization methods to perform accurate localization when one single dipole is simulated should be considered as “necessary (although not sufficient) condition for correct localization in general” (Pascual-Marqui, 1999). This ability may be assessed using the columns of the resolution matrix, originally proposed by Backus and Gilbert (1970). When a linear solution of the form $\hat{\mathbf{J}} = \mathbf{A} \cdot \mathbf{M}$ is available, the resolution matrix $\mathbf{R} = \mathbf{A} \cdot \mathbf{G}$ is a very interesting method to study the properties of distributed source localization methods before we get any data. Whereas the rows of \mathbf{R} generally refer to resolution kernels and the columns to point spread functions, many figures of merits derived from \mathbf{R} were proposed to compare distributed source localization methods, as for instance feature position and bias dipole localization (Grave de Peralta Menendez et al., 1996), source identifiability and source visibility (Grave de Peralta Menendez and Gonzales Andino, 1998) or cross-talk (Liu et al., 2002). However, such validation

metrics cannot be used to evaluate MEM-based methods as they do not provide a linear solution. Moreover, the resolution matrix does not assess the behavior of the methods when numerous or extended sources are simulated. Despite the superposition principle, we expect that more spurious sources would be generated when more than one dipolar source is simulated (e.g., spatially extended sources, combination of sources). We actually observed such spurious sources on our simulated data, and it was also suggested by Hämäläinen in his comment in Pascual-Marqui (1995). Whereas some authors proposed localization methods to quantitatively estimate the spatial extent (David and Garnero, 2002; Kincses et al., 2003), no validation metric to evaluate detection accuracy of distributed source localization methods with extended sources has yet been proposed.

Although the resolution matrix may inform us on the regions of the brain where one could expect more or less good localization before acquiring any data, we proposed here a complementary index to evaluate the whole process of source localization and notably the global detection accuracy of the methods, when realistic source configurations are simulated. To assess detection accuracy, we used the index AUC derived from ROC analysis of the normalized energy of the reconstructed activity, as suggested by Darvas et al. (2004). In this study, we proposed an adaptation of the standard AUC to circumvent bias inherent to application of ROC analysis with distributed sources localization: (1) the number of estimated or simulated inactive sources was much greater than the number of active sources, (2) we should provide an unbiased estimation of the false positive rate, (3) we should have enough statistic (i.e., enough points) to estimate the ROC parameters and (4) we should take into account the spatial resolution of the localization methods. Our criterion AUC was more appropriate than the ROC analysis proposed by Darvas et al. (2004) as we randomly selected as many fictive sources as the number of simulated active sources. To provide an unbiased estimation of the false positive rate, those fictive sources were either drawn within a close neighborhood of the simulated source ($\text{AUC}_{\text{close}}$) or within local maxima of activity distributed far from the source (AUC_{far}). The proposed index AUC, defined as the mean of $\text{AUC}_{\text{close}}$ and AUC_{far} , was thus sensitive to the ability of the method to accurately focalize the reconstructed activity around the source and to the generation of spurious sources far from the simulated source. Spatial resolution of the localization methods was taken into account for very focal sources by artificially increasing the area of theoretically active sources Θ_a before the ROC analysis. This improvement allowed us to consider that finding a source 1 cm away from the simulated one was still successful. Moreover, it ensures having enough statistic ($p_a \approx 24$ sources) for the estimation of the ROC parameters for the most focal sources. Finally, AUC was estimated as the mean AUC over 50 independent random samplings of the fictive sources (Θ_f). The dispersion of AUC around its mean was relatively small (results not shown), suggesting that our index provides a robust estimation of detection accuracy, not sensitive to one particular choice of Θ_f .

Comparison of localization methods

The proposed validation framework, including realistic spike simulations and original validation metrics, allowed us to better understand the behavior of six localization methods in realistic conditions and to tune some parameters of the methods. The index of detection accuracy AUC was in good agreement with the visual

interpretation of the results (see Figs. 3–5). We observed that, using a discrete Laplacian on a circunvoluted surface with fixed source orientations, numerical instabilities may occur, leading to complete mislocalizations. This issue is illustrated in Fig. 3 and was confirmed by our evaluation as LORETA1 was the least robust method in terms of AUC values, showing the highest variability. As suggested by David and Garnero (2002), we added a minimum norm term to the discrete Laplacian term in order to avoid these instabilities. The resulting method, LORETA2, was among the most accurate methods in our evaluation context, notably when detecting widespread sources, which was not surprising as such source configurations were in agreement with the assumption of maximum spatial smoothness. We may assume that most of those instabilities are due to the use of a discrete Laplacian on a closed circunvoluted surface. We notably observed that, when segmentation errors generated disconnections along the cortical surface, the occurrence of numerical instabilities increased. Applying the Laplacian constraint on a closed and disconnected part of the cortical surface may be completely misleading as the method may consider this area independently from the rest of the brain. One should thus prefer segmentation techniques preserving the topology of the cortical surface (Mangin, 1995), i.e., generating a closed surface following the circunvoluted geometry of the cortex and connecting both hemispheres together. Even if verifying such an assumption would require a specific study, the spatial connectivity of the cortical surface should be carefully checked before using “cortical” LORETA.

Validation using realistic simulated data is also an ideal framework to tune parameters of a new method for a specific context. We provide an evaluation of a promising method recently proposed by Amblard et al. (2004), the maximum entropy on the mean (MEM), that proved to be as accurate as LORETA2 in many configurations. However, the MEM approach is highly sensitive to the initialization of the reference distribution $d\mu$ that encompasses our prior knowledge of the current distribution. Amblard et al. (2004) proposed an interesting and flexible mixture model to define $d\mu$, but the sensitivity of the method to many parameters of such a model (cf. K , α_k , μ_k and Σ_k) is still an open question and will require further investigations. Here, we have studied the effect of the initialization of the mean activity within each cortical parcel (μ_k). Results showed that a better initialization of the mean activity within each parcel using the MN solution improved performance. Although the resulting MEM2 approach was among the most accurate methods of our study, it underestimated the amplitude of the reconstructed activity as the MN-based methods (cf. MSE and DF close to 1). This is certainly due to bias in the initialization of μ_k using MN, and we believe that we could achieve a more accurate estimation of the current amplitude using the iterative version of the MEM approach (Amblard et al., 2004). This point is currently under investigation.

We have shown that none of the methods were able to accurately localize mesial sources within the temporo-mesial and fronto-mesial areas. This observation may be explained by the fact that none of the evaluated methods, except the MSP, used a normalization of the lead field \mathbf{G} , so that superficial solutions were generally preferred (Fuchs et al., 1999). On the other hand, we found no clear effect of spatial extent on detection accuracy. At a moderate noise level (SNR = 6), LORETA2 and MEM2 were the most accurate methods (higher AUC), whereas LORETA-based and MEM-based methods showed the weakest results in terms of spurious sources (higher D_{\min}). Simulations involving pairs of

synchronous sources confirmed those trends. Similar findings were already reported in the literature for LORETA. The occurrence of spurious sources was not specifically studied by Pascual-Marqui (Pascual-Marqui, 1995, 1999; Pascual-Marqui et al., 2002), whereas some authors observed their occurrence using LORETA on a 3D grid (Michel et al., 2004b or Hämäläinen comments in Pascual-Marqui, 1995) or on a cortical surface (Trujillo-Barreto et al., 2004; Yamashita et al., 2004).

On the other hand, the APM maps estimated using the multivariate source prelocalization (Mattout et al., 2005) located very accurately the global maximum of activity in almost all configurations at SNR = 6 (see Tables 3 and 5), even in temporo-mesial and fronto-mesial areas. This multivariate analysis showed that the signatures of those mesial sources on the sensors (cf. fIBF), characterized by their specific source orientations, were highly correlated with the data, suggesting that sources were located there with high probability. Similar methods exploiting a singular value decomposition of the normalized lead field followed by a projection of the data on those principal components were already proposed, see for instance Fuchs et al. (1998) and EPIFOCUS (Grave de Peralta Menendez et al., 2001). MSP may be seen as a multivariate extension of those approaches. Similarly to the WMN method evaluated here, Fuchs et al. (1998) also used these projections for the purpose of the regularization within a weighted minimum norm framework but did not evaluate this method on distributed source models. EPIFOCUS (Grave de Peralta Menendez et al., 2001) showed very promising results in accurately localizing the maximum of the distribution in the context of interictal spikes localization (Michel et al., 2004a). However, the MSP was never evaluated in the context of extended sources, and APM as well as WMN seemed not really appropriate to recover large spatial extents (cf. lower AUC values). Nevertheless, as a preprocessing step, MSP showed high sensitivity and low specificity and brought useful information to initialize for example the MEM-based methods.

Increasing the noise level resulted in a decrease of detection accuracy (lower AUC) and in the generation of more spurious sources (higher D_{\min}) for most of the localization methods, except for MEM2. Noise particularly affects LORETA1, LORETA2 and APM which generate more spurious sources at SNR = 2. On the other hand, MEM2 was clearly the method the least sensitive to the level of noise. Whereas the variance of the noise σ_E^2 was estimated on the average background activity for MEM-based approaches, the level of noise was taken into account thanks to the heuristic approach of the “L curve” (Hansen, 1992) for the Bayesian methods. Moreover, the occurrence of spurious sources is certainly due to the fact that all the evaluated methods assumed the noise to be temporally and spatially uncorrelated, which is not the case in real data and in our realistic simulations. Methods derived from the Bayesian framework could certainly take benefit of a better estimation of the covariance of the noise and of the hyperparameter of regularization α to achieve lower sensitivity to noise and to generate less spurious sources. Within the Bayesian framework, we recently proposed a maximum a posteriori estimate of the regularization hyperparameter α (Daunizeau et al., 2005), allowing moreover the assessment of the relevance of any prior from the EEG data. We plan to use such a method to compare the reliability of different estimates $\hat{\mathbf{J}}$ corresponding to different priors, without any gold standard.

As suggested by Michel et al. (2004b), one should always take into account the results from different localization methods when

analyzing real data as converging evidence from different studies is needed. Following our results, one could recommend that, when analyzing real interictal spikes, LORETA2 and MEM2 should be compared, and the results should be concordant with the maximum of activity observed using APM coefficients in order to reject the influence of potential spurious sources and LORETA instabilities.

Conclusion and perspectives

We have proposed a validation framework using realistic simulations of interictal spikes and an original validation metric based on ROC curve analysis to assess the detection accuracy. By measuring areas under the ROC curve, we estimated a global behavior of a source localization method when recovering extended sources. To quantitatively estimate the spatial extent of the detected sources, we could use the ROC curves to threshold the normalized reconstructed energy by choosing the best trade-off between sensitivity and specificity. A better estimation of the false positive rate on real data could be obtained by using permutation tests (Darvas et al., 2004).

Six localization methods were evaluated here: two MN-based, two LORETA-based and two MEM-based approaches. This evaluation showed that distributed source localization methods were able to accurately recover sources likely to occur in epilepsy with a large range of spatial extents (from 10 mm² to 40 cm²), with the exception of sources in the temporo-mesial or fronto-mesial regions. Results showed that LORETA2 and MEM2 approaches provide the best detection accuracy. However, to detect possible spurious sources or to reject errors due to numerical instabilities, both methods should be compared with the maximum of activity detected using the MSP. Concerning the perspectives of the work, the proposed validation framework will allow adding the temporal dimension. We plan to evaluate the ability of the methods to recover accurately the time course of a spike generated by extended sources, as well as the propagation of the spike between several locations. Our results suggest that we should combine different source localization methods to provide the most accurate detection, and Bayesian model averaging as proposed by Trujillo-Barreto et al. (2004) seems to be a very promising framework to do so.

Acknowledgments

This work was supported by a post-doctoral fellowship (C. Grova) provided by the “Institut National pour la Recherche en Informatique et Automatique” (INRIA) in 2003 and by funds from the Canadian Institutes of Health Research (grant MOP-10189).

Appendix A. Bayesian framework

The Bayesian framework allows the estimation of the a posteriori distribution of \mathbf{J} given the data \mathbf{M} as a function of the data likelihood $Prob(\mathbf{M}|\mathbf{J}; \sigma_E)$ and of the prior $Prob(\mathbf{J}|\sigma_J)$, which represents the knowledge we may have on the distribution of \mathbf{J} (Baillet et al., 2001). σ_E and σ_J are two hyperparameters respectively associated to the variance of the noise \mathbf{E} and to the variance of the sources activity \mathbf{J} . Under Gaussian assumptions

regarding the distributions of \mathbf{J} and \mathbf{E} , with covariance matrix $Var[\mathbf{J}]$ given by $\sigma_J^2 (\mathbf{W} \cdot \mathbf{W}^T)^{-1}$ and $Var[\mathbf{E}]$ of the form $\sigma_E^2 \mathbf{I}_n$, the maximum a posteriori (MAP) estimator $\hat{\mathbf{J}}_{MAP}$ is obtained by minimizing the following cost function:

$$\hat{\mathbf{J}}_{MAP} = \operatorname{argmin}_{\mathbf{J}} \left[\|\mathbf{M} - \mathbf{G} \cdot \mathbf{J}\|^2 + \alpha \|\mathbf{W} \cdot \mathbf{J}\|^2 \right] \text{ with } \alpha = \frac{\sigma_E^2}{\sigma_J^2} \quad (\text{A.1})$$

This solution of the inverse problem is then given by:

$$\hat{\mathbf{J}}_{MAP} = (\mathbf{G}^T \cdot \mathbf{G} + \alpha \mathbf{W}^T \cdot \mathbf{W})^{-1} \cdot \mathbf{G}^T \cdot \mathbf{M} \quad (\text{A.2})$$

Appendix B. Maximum entropy on the mean (MEM)

Estimating \mathbf{J} as a realization of a random multivariate variable requires estimating its distribution $d\rho(\mathbf{j}) = \text{Prob}(\mathbf{J} = \mathbf{j})$. A reference model is introduced by writing the solution of the form $d\rho(\mathbf{j}) = f(\mathbf{j}) d\mu(\mathbf{j})$, where the reference distribution $d\mu$ expresses some prior information on \mathbf{J} and $f(\mathbf{j})$ is a density to be found such that it explains the data in average (noise being zero mean):

$$\mathbf{M} = \int \mathbf{G} \cdot \mathbf{j} f(\mathbf{j}) d\mu(\mathbf{j}) \quad (\text{B.1})$$

It is found that the MEM solution $d\rho$ is the density law in the exponential family of the reference law $d\mu$, from which the estimate is found to be (Amblard et al., 2004):

$$\hat{\mathbf{J}}_{MEM} = \nabla F_{\mu}(\xi)_{|\xi = \mathbf{G}^T \tilde{\lambda}} \quad (\text{B.2})$$

$$\text{with } F_{\mu}(\xi) = \log \int \exp(\xi^T \mathbf{j}) d\mu(\mathbf{j}) \quad (\text{B.3})$$

$$\text{and } \tilde{\lambda} = \operatorname{argmax}_{\lambda} \left(\lambda^T \mathbf{M} - F_{\mu}(\mathbf{G}^T \lambda) - \frac{\sigma_E^2}{2} \lambda^T \cdot \lambda \right) \quad (\text{B.4})$$

where the “free energy” F_{μ} is a normalization term that ensures $d\rho(\mathbf{j})$ to be a probability distribution and $\tilde{\lambda}$ is an n -dimensional vector obtained from a non-linear optimization in an n -dimensional space and σ_E^2 is the variance of the noise as defined in Eq. (A.1).

References

- Alarcon G., Guy C., Walker S., Elwes R., Polkey C., 1994. Intracerebral propagation of interictal activity in partial epilepsy: implications for source localization. *J. Neurol. Neurosurg. Psychiatry* 57, 435–449.
- Al-Asmi A., Bénar C.G., Gross D., Aghakhani Y., Andermann F., Pike B., Dubeau F., Gotman J., 2003. fMRI activation in continuous and spike-triggered EEG-fMRI studies of epileptic spikes. *Epilepsia* 44 (10), 1328–1339.
- Amblard C., Lapalme E., Lina J.M., 2004. Biomagnetic cortical sources reconstruction by maximum entropy on the mean. *IEEE Trans. Biomed. Eng.* 51 (3), 427–442.
- Backus G., Gilbert F., 1970. Uniqueness in the inversion of inaccurate gross earth data. *Philos. Trans. R. Soc. Lond., A* 266, 123–192.
- Baillet S., Mosher J., Leahy R., 2001. Electromagnetic brain mapping. *IEEE Signal Process. Mag.*, 14–30.
- Baumgartner C., 2004. Controversies in clinical neurophysiology. MEG is superior to EEG in the localization of interictal epileptiform activity: *Con. Clin. Neurophysiol.* 115 (5), 1010–1020.

- Bénar C.G., Gunn R., Grova C.G., Champagne B., Gotman J., 2005. Statistical maps for EEG dipolar source localization. *IEEE Trans. Biomed. Eng.* 52 (3), 401–413.
- Bjorkstrom, A. (2000). Ridge regression and inverse problem. Technical report, Research reports in Mathematical Statistics, Department of Mathematics Stockholm University. URL:<http://www.math.su.se/matstat/reports/serica/2000/rep5/report.pdf>.
- Chassoux F., Semah F., Bouillere V., Landre E., Devaux B., Turak B., Nataf F., Roux F., 2004. Metabolic changes and electroclinical patterns in mesiotemporal lobe epilepsy: a correlative study. *Brain* 127 (1), 1–11.
- Chauvel P., Vignal J., Biraben A., Badier J.M., Scarabin J.M., 1996. In: Pawlik, G., Stefan, H. (Eds.), *Multimethodological Assessment of the Epileptic Forms. Chapter Stereoecephalography*. Springer Verlag, New York, pp. 80–108.
- Clarke C., Janday B., 1989. The solution of biomagnetic inverse problem by maximum statistical entropy. *Inverse Probl.* 5, 483–500.
- Cooper R., Winter A.L., Crow H.J., Walter W.G., 1965. Comparison of subcortical, cortical and scalp activity using chronically indwelling electrodes in man. *Electroencephalogr. Clin. Neurophysiol.* 18, 217–228.
- Dale A., Sereno M., 1993. Improved localization of cortical activity by combining EEG and MEG with MRI surface reconstruction: a linear approach. *J. Cogn. Neurosci.* 5, 162–176.
- Darvas F., Pantazis D., Kucukaltun-Yildirim E., Leahy R., 2004. Mapping human brain function with MEG and EEG: methods and validation. *NeuroImage* 23, S289–S299.
- Daunizeau J., Mattout J., Goulard B., Lina J.M., Benali H., 2004. Data-driven cortex parcelling: a regularization tool for the EEG/MEG inverse problem. *Proceedings of the IEEE International Conference on Image Processing (ISBI 2004)*, pp. 1343–1346.
- Daunizeau J., Grova C., Mattout J., Marrelec G., Clonda D., Goulard B., Péligrini-Issac M., Lina J.M., Benali H., 2005. Assessing the relevance of fMRI-based prior in the EEG inverse problem: a Bayesian model comparison approach. *IEEE Trans. Signal Process.* 53 (9), 3461–3472 (Special issue on data fusion in neuroimaging).
- David O., Garnero L., 2002. Time-coherent expansion of MEG/EEG cortical sources. *NeuroImage* 17, 1277–1289.
- Devous M., Thisted R., Morgan G., Leroy R., Rowe C., 1998. SPECT brain imaging in epilepsy: a meta-analysis. *J. Nucl. Med.* 39 (2), 285–293.
- Ebersole J., 1997. Defining epileptogenic foci: past, present, future. *J. Clin. Neurophysiol.* 14 (6), 470–483.
- Efron B., Tibshirani R., 1993. An introduction to the bootstrap. Number 57 in *Monographs on Statistics and Applied Probability*. Chapman and Hall/CRC.
- Fuchs M., Wagner M., Wischmann H., Kohler T., Theissen A., Drenkhahn R., Buchner H., 1998. Improving source reconstructions by combining bioelectric and biomagnetic data. *Electroencephalogr. Clin. Neurophysiol.* 107, 93–111.
- Fuchs M., Wagner M., Kohler T., Wischmann H.-A., 1999. Linear and nonlinear current density reconstructions. *J. Clin. Neurophysiol.* 16 (3), 267–295.
- Gorodnitsky I., George J., Rao B., 1995. Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm. *Electroencephalogr. Clin. Neurophysiol.* 95, 231–251.
- Gotman J., Bénar C., Dubeau F., 2004. Combining EEG and fMRI in epilepsy: methodological challenges and clinical results. *J. Clin. Neurophysiol.* 21 (4), 229–240.
- Grave de Peralta Menendez R., Gonzales Andino S., 1998. A critical analysis of linear inverse solutions to the neuroelectromagnetic inverse problem. *IEEE Trans. Biomed. Eng.* 45 (4), 440–448.
- Grave de Peralta Menendez R., Gonzales Andino S., Lutkenhoner B., 1996. Figures of merit to compare distributed linear inverse solutions. *Brain Topogr.* 9 (2), 117–124.
- Grave de Peralta Menendez R., Gonzales Andino S., Lantz G., Michel C., Landis T., 2001. Noninvasive localization of electromagnetic epileptic activity, I. Method descriptions and simulations. *Brain Topogr.* 14 (2), 131–137.
- Hamalainen M., Ilmoniemi R., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* 32, 35–42.
- Hansen P., 1992. Analysis of ill-posed problems by means of the L-curve. *SIAM Rev.* 34, 561–580.
- Im C., An K., Jung H., Kwon H., Lee Y., 2003. Assessment criteria for MEG/EEG cortical patch tests. *Phys. Med. Biol.* 48, 2561–2573.
- Jannin P., Fitzpatrick J., Hawkes D., Pennec X., Shahidi R., Vannier M., 2002. Editorial: validation of medical image processing in image-guided therapy. *IEEE Trans. Med. Imag.* 21 (11), 1445–1449.
- Kincses W., Braun C., Kaiser S., Grodd W., Ackermann H., Mathiak K., 2003. Reconstruction of extended cortical sources for EEG and MEG based on a Monte-Carlo-Markov-Chain estimator. *Hum. Brain Mapp.* 18, 100–110.
- Kobayashi K., Yoshinaga H., Oka M., Ohtsuka Y., Gotman J., 2003. A simulation study of the error in dipole source localization for EEG spikes with a realistic head model. *Clin. Neurophysiol.* 114, 1069–1078.
- Koles Z., 1998. Trends in EEG source localization. *Electroencephalogr. Clin. Neurophysiol.* 106, 219–230.
- Liu A., Belliveau J., Dale A., 1998. Spatiotemporal imaging of human activity using functional MRI constrained magnetoencephalography data: Monte Carlo simulations. *Proc. Natl. Acad. Sci. U. S. A.* 95, 8945–8950.
- Liu A., Dale A., Belliveau J., 2002. Monte Carlo simulation studies of EEG and MEG localization accuracy. *Hum. Brain Mapp.* 16, 47–62.
- Luders H., Awad I., 1992. In: Luders, H.O. (Ed.), *Epilepsy Surgery: Conceptual Considerations*. Raven Press Publishers, pp. 51–62.
- Mangin J.F., 1995. From 3D magnetic resonance imaging to structural representation of the cortex topography using topology preserving deformations. *J. Math. Imaging Vis.* 5, 297–318.
- Mattout J., Péligrini-Issac M., Garnero L., Benali H., 2005. Multivariate source prelocalization: use of functionally informed basis functions for better conditioning the MEG inverse problem. *NeuroImage* 26 (2), 356–373.
- Merlet I., Gotman J., 1999. Reliability of dipole models of epileptic spikes. *Clin. Neurophysiol.* 110 (6), 1013–1028.
- Metz C., 1986. ROC methodology in radiologic imaging. *Invest. Radiol.* 21 (9), 720–732.
- Michel C., Lantz G., Spinelli L., Grave de Peralta R., Landis T., Seck M., 2004a. 128-channel EEG source imaging in epilepsy: clinical yield and localization precision. *J. Clin. Neurophysiol.* 21 (2), 71–83.
- Michel C., Murray M., Lantz G., Gonzales S., Spinelli L., Grave de Peralta R., 2004b. EEG source imaging. *Clin. Neurophysiol.* 115, 2195–2222.
- Morozov V., 1968. The error principle in the solution of operators equations by the regularization method. *USSR Comp. Math. Phys.* 28, 69–80.
- Mosher J., Leahy R., 1999. Source localization using recursively applied and projected (RAP) MUSIC. *IEEE Trans. Signal Process.* 47 (2), 332–340.
- Mosher J., Leahy R., Lewis P., 1999. EEG and MEG: forward solutions for inverse methods. *IEEE Trans. Biomed. Eng.* 46, 245–259.
- Otsu N., 1979. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* 9 (1), 62–69.
- Pascual-Marqui R., 1995. Reply to comments by Hamalainen, Ilmoniemi and Nunez. In *Source Localization: continuing discussion of the inverse problem*. *ISBET Newsl.* (6), 16–28.
- Pascual-Marqui R., 1999. Review of methods for solving the EEG inverse problem. *Int. J. Bioelectromagn.* 1 (1), 75–86.
- Pascual-Marqui R., 2002. Standardized low resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24D, 5–12.
- Pascual-Marqui R., Michel C., Lehman D., 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.* 18, 49–65.

- Pascual-Marqui R., Esslen M., Kochi K., Lehmann D., 2002. Functional imaging with low resolution brain electromagnetic tomography (LORETA): review, new comparisons, and new validation. *Jpn. J. Clin. Neurophysiol.* 30, 81–94.
- Scherg M., Von Cramon D., 1986. Evoked dipole source potentials of the human auditory cortex. *Electroencephalogr. Clin. Neurophysiol.* 65, 344–360.
- Schmidt D., George J., Wood C., 1999. Bayesian inference applied to the electromagnetic inverse problem. *Hum. Brain Mapp.* 7, 195–212.
- Sekihara K., Nagarajan S., Poeppel D., Marantz A., Miyashita Y., 2001. Reconstructing spatio-temporal activities of neural sources using an MEG vector beamformer technique. *IEEE Trans. Biomed. Eng.* 48 (7), 760–771.
- Speckmann E.-J., Elger C., Altrup U., 2004. Neurophysiologic basis of the electroencephalogram. In: Niedermeyer, E., Lopes Da Silva, F. (Eds.), *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 5th ed. Lippincott, Williams and Wilkins, Baltimore, pp. 149–163 (chapter 12).
- Spencer S., 2002. Neural networks in human epilepsy: evidence of and implications for treatment. *Epilepsia* 43 (3), 219–227.
- Stephen J., Aine C., Ranken D., Hudson D., Shih J., 2003. Multidipole analysis of simulated epileptic spikes with real background activity. *J. Clin. Neurophysiol.* 20 (1), 1–16.
- Tikhonov A., Arsenin V., 1977. *Solutions of Ill-Posed Problems*. W.H. Winston, Washington, DC. (English Translation by F. John).
- Trujillo-Barreto N., Aubert-Vazquez E., Valdes-Sosa P., 2004. Bayesian model averaging in EEG/MEG imaging. *NeuroImage* 21, 1300–1319.
- Wagner M., Fuchs M., Drenckhahn R., Wischmann H.-A., Köhler T., Theißen A., 1997. Automatic generation of BEM and FEM meshes. *NeuroImage* 5, 389.
- Yamashita O., Galkas A., Ozaki T., Biscay R., Valdes-Sosa P., 2004. Recursive penalized least squares solution for dynamical inverse problems of EEG generation. *Hum. Brain Mapp.* 21, 221–235.