# Fast accurate MEG source localization using a multilayer perceptron trained with real brain noise

**Sung Chan Jun, Barak A. Pearlmutter, Guido Nolte**

Dept. of Computer Science, University of New Mexico, Albuquerque, NM 87131, U.S.A.

E-mail: junsc@cs.unm.edu, bap@cs.unm.edu, nolte@cs.unm.edu

**Abstract.** Iterative gradient methods like Levenberg-Marquardt (LM) are in widespread use for source localization from electroencephalographic (EEG) and magnetoencephalographic (MEG) signals. Unfortunately LM depends sensitively on the initial guess, necessitating repeated runs. This, combined with LM's high per-step cost, makes its computational burden quite high. To reduce this burden, we trained a multilayer perceptron (MLP) as a real-time localizer. We used an analytical model of quasistatic electromagnetic propagation through a spherical head to map randomly chosen dipoles to sensor activities according to the sensor geometry of a 4D Neuroimaging Neuromag-122 MEG system, and trained a MLP to invert this mapping in the absence of noise or in the presence of various sorts of noise such as white Gaussian noise, correlated noise, or real brain noise. A MLP structure was chosen to trade off computation and accuracy. This MLP was trained four times, with each type of noise. We measured the effects of initial guesses on LM performance, which motivated a hybrid MLP-start-LM method, in which the trained MLP initializes LM. We also compared the localization performance of LM, MLPs, and hybrid MLP-start-LMs for realistic brain signals. Trained MLPs are much faster than other methods, while the hybrid MLP-start-LMs are faster and more accurate than fixed-4-start-LM. In particular, the hybrid MLP-start-LM initialized by a MLP trained with the real brain noise dataset is 60 times faster and is comparable in accuracy to random-20-start-LM, and this hybrid system (localization error: 0.28 cm, computation time: 36 ms) shows almost as good performance as optimal-1-start-LM (localization error: 0.23 cm, computation time: 22 ms), which initializes LM with the correct dipole location. MLPs trained with noise perform better than the MLP trained without noise, and the MLP trained with real brain noise is almost as good an initial guessor for LM as the correct dipole location.

Physics in Medicine and Biology, *47*(14):2547–2560, June 21 2002.

## 1. Introduction

Source localization using EEG and MEG signals is important in medical diagnosis of conditions like epilepsy, in surgical planning, and in neuroscience research. The goal of this localization is to identify electrically active brain regions, which emit signals measured by EEG and MEG. There are a number of popular localization methods (Hämäläinen et al., 1993) most of which assume a dipolar source. The approach taken by most methods to solve the dipole source localization problem is:

- Calculate the sensor activations $\mathbf{B}_c(\mathbf{x})$ for dipole parameters $\mathbf{x}$ through a forward model.
- Calculate the cost function $c(\mathbf{x})$, a measure of the difference between the measured sensor activations $\mathbf{B}_m$ and the calculated sensor activations,
$$c(\mathbf{x}) = |\mathbf{B}_c(\mathbf{x}) - \mathbf{B}_m|^2.$$
- Adjust the dipole parameters $\mathbf{x}$ in order to reduce $c(\mathbf{x})$.
- Repeat to convergence.

Gradient methods, which calculate the gradient $\nabla_{\mathbf{x}} c(\mathbf{x})$ in choosing a change to $\mathbf{x}$, seem superior in terms of accuracy and computational burden (Press et al., 1988).

However, gradient methods require both a differentiable forward model and an initial guess. As we shall see, the efficiency and accuracy of the most popular gradient method for this problem, LM (Levenberg, 1944; Marquardt, 1963), depends sensitively on the initial guess. There is therefore motivation to build faster and more accurate source localizers. This is particularly important for our real time MEG brain-computer interface system, as we need to localize BSS-separated components in real time.

Since it is easy to create synthetic data consisting of pairs of corresponding dipole locations and sensor signals, it is tempting to train a universal approximator to solve the inverse problem directly, *i.e.* to map sensor signals directly to the dipole location. The multilayer perceptron (MLP) of Rumelhart et al. (1986) has been popular for this purpose. MLPs were first used for EEG dipole source localization and presented as feasible source localizers by Abeyratne et al. (1991), and a MLP structure composed of six separate networks was later used for EEG dipole localization by Zhang et al. (1998). Kinouchi et al. (1996) first used MLPs for MEG source localization by training on a noise-free dataset of near-surface dipoles, and Yuasa et al. (1998) studied the two-dipole case for EEG dipole source localization while restricting each source dipole to a small region. Hoey et al. (2000) investigated EEG measurements for both spherical and realistic head models, trained on a randomly generated noise-free dataset, and presented a comparison between a MLP and an iterative method for localization with noisy signals at three fixed dipole locations. Sun and Sclabassi (2000) adapted a MLP to calculate forward EEG solutions for a spheroidal head model from simple EEG solutions for a spherical head model. Recently, Kamijo et al. (2001) proposed an integrated approach to EEG dipole source localization in which a MLP trained with noise-free data is used as an initializer for Powell's method.

The human skull phantom study of Leahy et al. (1998) shows that the fitted spherical head model for MEG localization is slightly inferior in accuracy to the realistic head model numerically calculated by a boundary element method (BEM). In forward calculation, a spherical head model has some advantages: it is more easily implemented, a forward calculation through the model is much faster, and the model parameters can be fit to a subject much more easily. Despite its inferiority in terms of localization accuracy, we use a spherical head model in this work.‡ We train a MLP to localize dipoles from noisy MEG measurements for a spherical head, and measure the efficacy of the resulting network under a variety of conditions. In Section 2 the forward model and noise model are explained in detail. Section 3 explores the tradeoff between accuracy and computation time in terms of MLP size, and present MLP training and learning curves for various datasets. Section 3 continues by exploring how S/N and other system parameters affect localization accuracy. The effects of initial guesses on the performance of LM are simulated in Section 4, and Section 5 compares the performance of LM, MLP and hybrid MLP-start-LM. We conclude that MLPs can serve as real-time MEG localizers, MLPs should be trained with noise, and the MLP trained with real brain noise is almost as good an initial guessor for LM as the correct dipole location.

## 2. Synthetic data

The synthetic data used in our experiments consisted of corresponding pairs of dipole locations and sensor activations, as generated by our forward model. Given a dipole location and a set of sensor activations, the minimum error dipole moment can be calculated analytically (see Section 4). Therefore, although the dipoles used in generating the data set have both location and moment, we discarded the moment in all the experiments below.

We made two datasets, one for training and the other for testing. Dipoles in the training and testing sets were drawn uniformly from truncated spherical regions, as shown in Figure 1. Their moments were drawn uniformly from vectors of strength $\leq 100$ nAm. The corresponding sensor activations were calculated by adding the results of a forward model and a noise model. To allow the network to interpolate rather than extrapolate, thus improving performance, the training set used dipoles from the larger region, while the test set contained only dipoles from the smaller inner region.

### *2.1. Forward model*

We used a standard analytic forward model of quasistatic electromagnetic propagation in a spherical head (Sarvas, 1987; Mosher et al., 1999), with the sensor geometry of a 4D

‡ This work could be easily extended to a more realistic forward model. One can expect that a more complex forward model leads to more local optima, and can therefore degrade the performance of gradient-based methods. However, it should not much affect the performance of a trained MLP. For this reason, we would expect a more sophisticated head model to give a comparatively greater advantage to the MLP-based approach advocated in this paper. For realistic volume conductors algorithms like seed-based Simplex are common. Using the MLP to get an initial guess does not require optimal accuracy, and we expect a spherical approximation for the training of the MLP to be sufficient for this application.

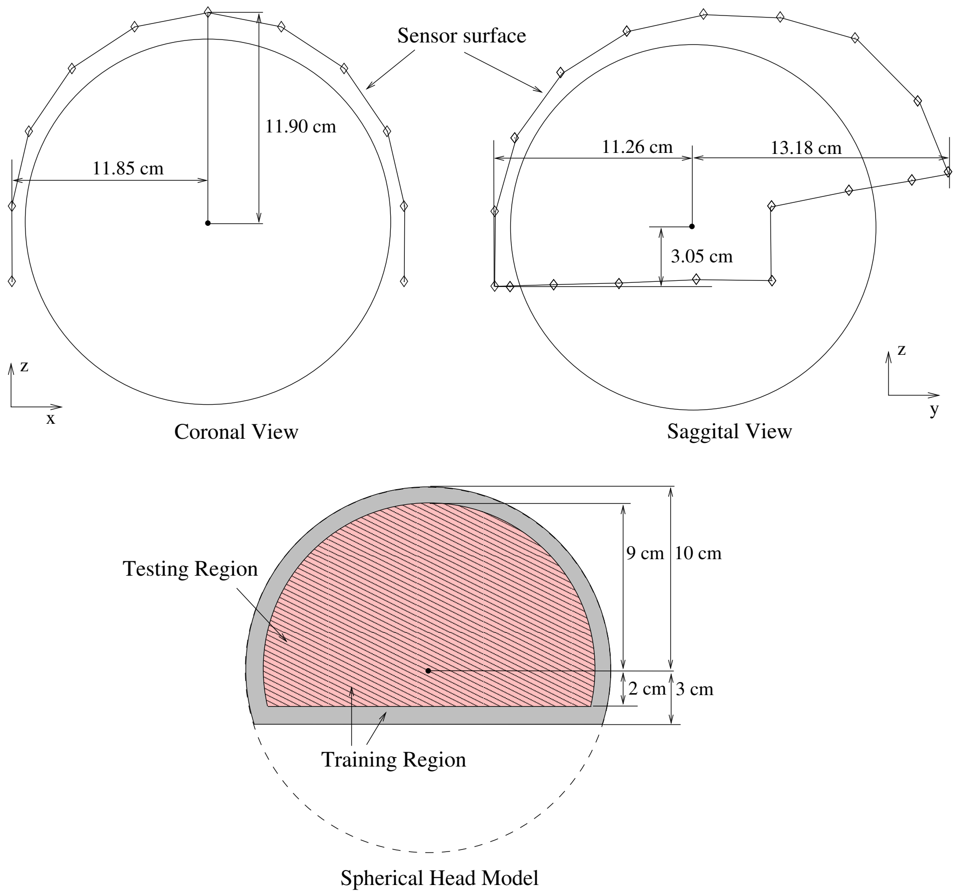


**Figure 1.** Sensor surface and spherical head model. Training and testing regions for a spherical head model. Diamonds denote sensors.

Neuroimaging Neuromag-122 gradiometer.

$$
\begin{aligned}
&B_s(\mathbf{x}, \mathbf{Q}) = \frac{\left[\mathbf{M}(\mathbf{x}, \mathbf{Q}; \mathbf{t})|_{\mathbf{t}=\mathbf{x}_s^1} - \mathbf{M}(\mathbf{x}, \mathbf{Q}; \mathbf{t})|_{\mathbf{t}=\mathbf{x}_s^2}\right] \cdot \mathbf{r}_s}{|\mathbf{x}_s^1 - \mathbf{x}_s^2|}, \quad s = 1, \cdots, 122, \\
&\mathbf{M}(\mathbf{x}, \mathbf{Q}; \mathbf{x}_s) = \frac{\mu_0}{4\pi} \frac{F\mathbf{Q} \times \mathbf{x} - (\mathbf{Q} \times \mathbf{x} \cdot \mathbf{x}_s)\nabla F_{\mathbf{x}}}{F^2}, \\
&F(\mathbf{x}, \mathbf{x}_s) = d(x_s d + {x_s}^2 - (\mathbf{x}_s \cdot \mathbf{x})) \\
&\nabla F_{\mathbf{x}}(\mathbf{x}, \mathbf{x}_s) = \left(\frac{d^2}{x_s} + \frac{(\mathbf{d} \cdot \mathbf{x}_s)}{d} + 2d + 2x_s\right)\mathbf{x}_s - \left(d + 2x_s + \frac{(\mathbf{d} \cdot \mathbf{x}_s)}{d}\right)\mathbf{x} \\
&\mathbf{d} = \mathbf{x}_s - \mathbf{x}, \quad d = |\mathbf{x}_s - \mathbf{x}|, \quad x_s = |\mathbf{x}_s|,
\end{aligned}
$$

where $\mathbf{x}$ and $\mathbf{Q}$ denote a source dipole location vector and a source dipole moment vector, respectively. The vectors $\mathbf{x}_s^1$ and $\mathbf{x}_s^2$ denote the positions of the centers of the first and second coils of $s$-th gradiometer sensor, and $\mathbf{r}_s$ denotes the orientation vector of the $s$-th sensor. $B_s(\mathbf{x}, \mathbf{Q})$ is the sensor activation of $s$-th sensor through the forward model, and $\mu_0$ is the permeability constant of air.

### *2.2. Noise model*

For single-trial data, the sensors in MEG systems have poor S/N ratios because MEG data is strongly contaminated not only by intrinsic sensor noise, but also by external fields, fields generated by various parts of the body (heart, eye muscles, retina), and signals from parts of the brain not under study. Blind source separation of MEG data can drastically improve the situation by segregating noise from signal (Vigário et al., 1998; Tang, Pearlmutter, Zibulevsky and Carter, 2000), and the sensor attenuation vectors of the BSS-separated components can be well localized to equivalent current dipoles (Tang, Phung, Pearlmutter and Christner, 2000; Tang et al., 2002). However, the recovered field maps can be quite noisy, and conventional localization techniques require manual interaction.

In order to compare the performance of various localizers, we need a dataset for which we know the ground truth, but which contains the sorts of noise encountered in actual MEG recordings. To this end, we created three noise processes with which to additively contaminate synthetic sensor readings (Kwon et al., 2000). These are: white Gaussian noise, correlated noise, and real brain noise. By using a variety of noise models, we achieve a rough measurement of the robustness of the system to a mismatch between the noise model used in training and the noise encountered in testing. The white Gaussian noise is generated by simply drawing a zero mean Gaussian-distributed random number for each sensor. Correlated noise is made using the method of Lütkenhöner (1994):

- Distribute 871 dipoles uniformly on a spherical surface, with dipole moments drawn from a zero-mean spherical Gaussian.
- Calculate a sensor activation through the analytic forward model for each dipole for each sensor and sum over all dipoles at each sensor.
- Scale the resultant sensor activation vector to yield a suitable RMS power.

Real brain noise was taken from MEG recordings during periods in which the brain region of interest in the experiment was quiescent. These signals were not averaged. The real brain noise has an RMS power $P^n$ of roughly 50–100 fT/cm. We measured the S/N ratio of a dataset using the ratios of the powers in the signal and the noise: S/N (in dB) $= 10 \log_{10} P^{\mathrm{s}}/P^{\mathrm{n}}$ where $P^{\mathrm{s}}$ is the RMS (square root of mean square) of the sensor readings from the dipole and $P^{\mathrm{n}}$ is the RMS power of the sensor readings from the noise.

The noisy datasets were made by adding noise to synthetic sensor activations generated by the forward model. Exemplars whose resulting S/N ratio was under $0$ dB were rejected. Real brain noise taken from MEG recordings was added without scaling, while the white Gaussian noise and the correlated noise were scaled to make the RMS power of the noise equal to that of the real brain noise.

## 3. Multilayer Perceptron

It is well known that a MLP with one or more hidden layers is a universal approximator (Hornik et al., 1989). As in Hoey et al. (2000), our experiments showed that a MLP with two

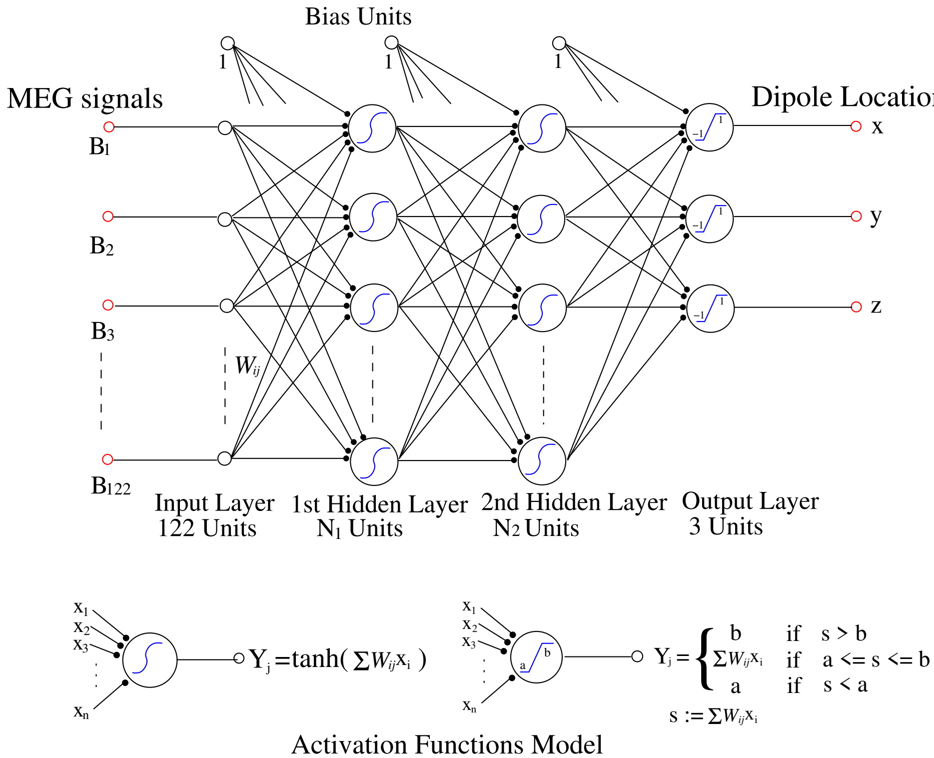


**Figure 2.** Training datasets were trained in this MLP structure.

hidden layers was superior to a MLP with one hidden layer, and for this reason we used a MLP with two hidden layers. The MLP charged with approximating the inverse mapping (Figure 2) had an input layer of 122 units, one for each sensor, two hidden layers with $N_1$ and $N_2$ units respectively, and an output layer of three units representing the dipole location $(x, y, z)$. The output units had linear activation functions, while the hidden units had hyperbolic tangent activation functions to accelerate training (LeCun et al., 1991). All units had bias inputs, adjacent layers were fully connected, and there were no cut-through connections.

Input data is usually preprocessed to improve the performance, and output data is scaled and translated into the dynamic range of the output unit activation function to avoid driving the weights of the network to infinity or driving the hidden neurons to saturation (Haykin, 1994). The 122 MEG sensor activations were scaled so that the RMS value was 0.5. Dipole location parameters were scaled to lie in the range $\pm 0.8$. The network weights were initialized with uniformly distributed random values between $\pm 0.1$. Backpropagation was used to calculate the gradient, and online stochastic gradient decent for the optimization. No momentum was used, and the learning rate was chosen empirically. Sometimes this procedure is referred to as *vanilla backpropagation.*

### *3.1. MLP structural optimization*

Beginning with intuitions drawn from the explorations of suitable numbers of hidden units by Hoey et al. (2000) for EEG localization, we empirically measured the tradeoff between

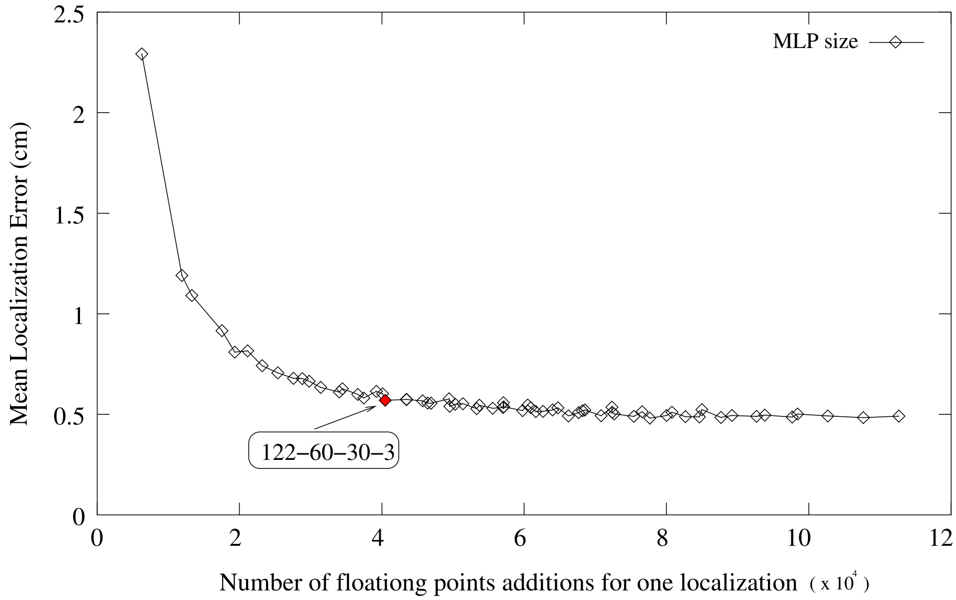


**Figure 3.** Mean localization error versus calculation time, as a parametric function of MLP structure. A noise-free synthetic dataset was used for training.

accuracy and computation time. Generalization was not a serious consideration, since training sets of arbitrary size could be easily constructed: our training sets ranged from 5 000–20 000, as circumstances dictated.

For practical reasons, we constrained our experiments to networks with no more than 110 hidden units in either hidden layer. We varied the number of hidden units in each layer from 10 to 110, in steps of 10, with $N_1 \geq N_2$. Each MLP was trained with a noise-free training dataset of 5 000 training exemplars, and the mean localization error for a noise-free training dataset of 5 000 was measured after 500 epochs, where in each epoch all exemplars in the training dataset are presented once. Training each network took up to two hours on an 800 MHz AMD Athlon computer. For each size, five runs were performed and the errors averaged. The calculation time was measured in terms of equivalent additions for a forward pass, *i.e.* a localization. The number of equivalent additions per multiplication and hyperbolic tangent were about 3 and 33 respectively, as measured on the above CPU. The equivalent floating points additions for a pass through the 122–$N_1$–$N_2$–3 MLP structure was therefore $(122N_1 + N_1N_2 + 3N_2) + 3(N_1N_2 + 123N_1 + 4N_2 + 3) + 33(N_1 + N_2)$.

Figure 3 shows average localization error for training *vs.* calculation time for a localization. Each point in the figure describes a different network architecture. Increasing the computation reduces the localization error. The accuracy levels off after a while, probably due to incomplete convergence of the network training. From this result we choose a MLP size of 122–60–30–3, as indicted in Figure 3. This is the most accurate MLP among those having computation time less than 40 000 additions, which takes less than one hour to train.

**Table 1.** Distribution of S/N for the 4 500 testing patterns

| S/N (dB) | Number of patterns | Frequency (%) |
|---|---|---|
| 0–1 | 892 | 19.82 |
| 1–2 | 806 | 17.91 |
| 2–3 | 824 | 18.31 |
| 3–4 | 627 | 13.93 |
| 4–5 | 538 | 11.96 |
| 5–6 | 381 | 8.47 |
| 6–7 | 230 | 5.11 |
| 7–8 | 118 | 2.62 |
| 8–9 | 53 | 1.18 |
| 9–10 | 23 | 0.51 |
| $>10$ | 8 | 0.18 |

### *3.2. Train of MLP and localization results*

The larger the training dataset, the better the generalization. In our work, we extracted real noise from an actual MEG recording for one subject, and the number of extracted noise patterns was therefore limited to about 25 000. For this reason we constructed four training datasets, each with 20 000 exemplars, differing only in the type of noise: none, white Gaussian, correlated, and real brain (Section 2.2). We always tested using the testing dataset consisting of 4 500 patterns contaminated by real brain noise. As described in Section 2, the S/N ratio was controlled by scaling additive noise so that the S/N ratio distributions for white Gaussian noise and correlated noise datasets matched the real brain noise dataset. The S/N ratio distributions for the testing dataset are shown in Table 1.

We used 500 epochs which took about four hours on an 800 MHz AMD Athlon to train a network of the selected architecture on a noisy dataset. Figure 4 shows the training and testing curves for our four cases: none, white Gaussian, correlated, and real brain noise. The noise-free dataset was optimized well, but even the noisiest dataset lead to less than 0.9 cm mean localization error. In terms of mean localization error for the testing dataset, the real brain noise is the most accurate, the white Gaussian noise was slightly more accurate than the correlated noise, and the noise-free was the worst. Obviously, the closer the training dataset was to the testing dataset, the better was the performance. The localization errors for 4 500 testing patterns of four trained MLPs, trained with various sorts of noise, are shown as a function of S/N in Figure 5. The MLP trained with real brain noise showed the best performance for low S/N ratio signals among the four systems, while all are comparable in accuracy at high S/N ratios. In conclusion, these results show that the MLP trained with a noisy dataset showed better performance than the MLP trained with the noise free dataset, while the MLP trained with real brain noise was the best.

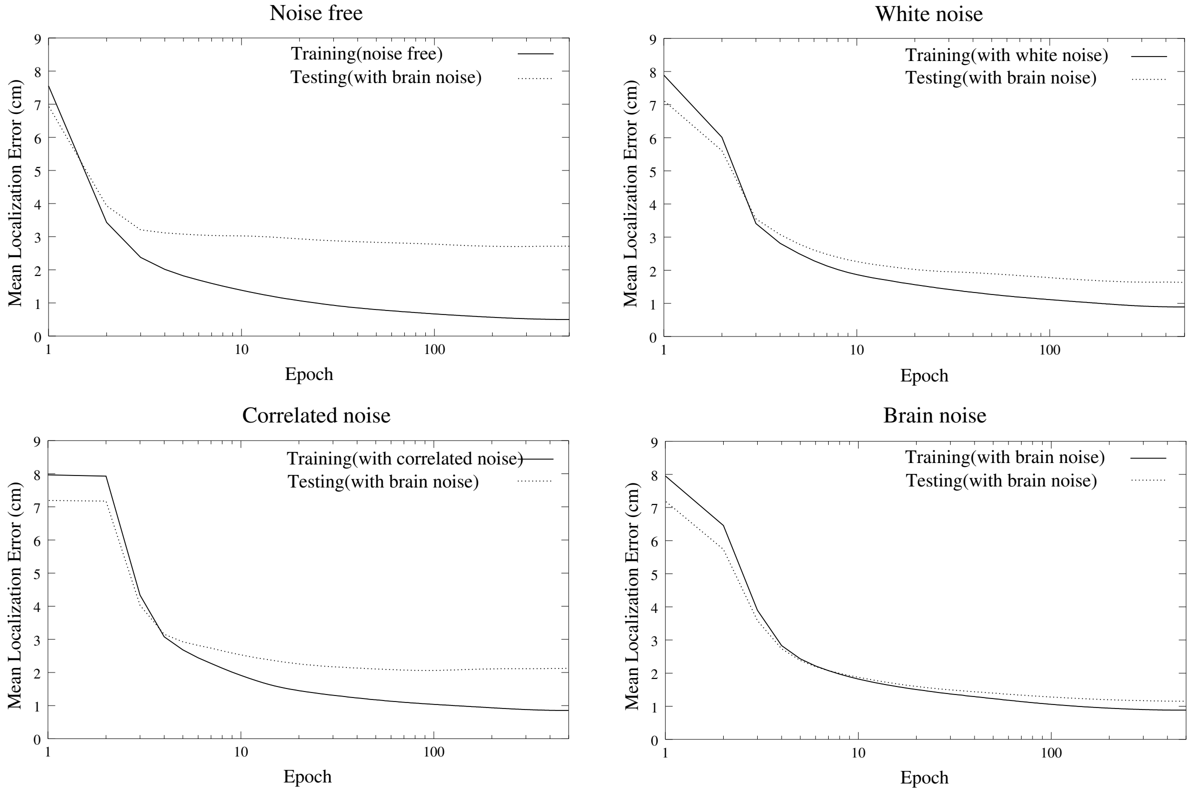


**Figure 4.** Mean localization error versus epoch for training of 20 000 exemplars with various type of noise: none, white Gaussian, correlated, and real brain. In all cases, testing used 4 500 patterns with real brain noise.

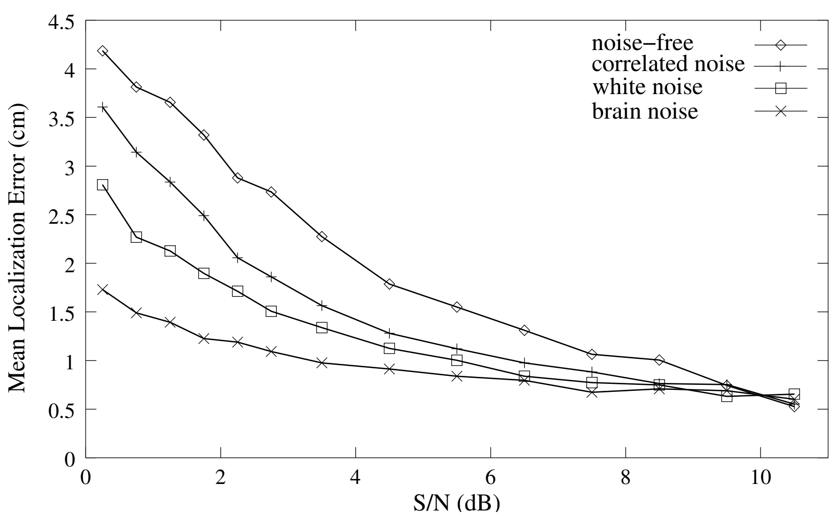


**Figure 5.** Mean localization error versus S/N for four MLPs trained with various training datasets: noise-free, white Gaussian noise, correlated noise, and real brain noise. Localization was performed for 4 500 real brain noisy signals.

### *3.3. Localization performance of MLP over various regions*

We investigated localization error distributions over various regions of interest. We considered two cross sections (coronal and saggital views), and these were divided into 36 and 28 regions, respectively, as shown in Figure 6. For each region, 100 real brain noise signals were constructed in the same way as the real brain noise dataset. A dipole localization was performed using the MLP trained with real brain noise and we calculated the average localization error for each region. Figure 6 shows the localization error distribution over the two cross sections. In general, dipoles closer to sensor surface were better localized.

## 4. Effect of the initial guess on LM Localization

To see how the initial guess affects the LM localizer, we measured the localization performance of LM as a function of the distance from the initial dipole location to the actual location. The initial guess was chosen randomly on a sphere of radius $d$ centered on the target. For each S/N ratio, 300 correlated noisy patterns were created. The S/N ranged from 0–11 dB, and the distance $d$ from 0–6 cm in steps of 1 cm. For each sensor activation and initial guess, LM finds the dipole parameters that minimize a quadratic function (Hämäläinen et al., 1993) of the difference between the predicted and input sensor activations,

$$c(\mathbf{x}) = ||\mathbf{B}(\mathbf{x}, \mathbf{Q}(\mathbf{x})) - \mathbf{B}_m||^2_{\Sigma^{-1}}$$

where $||\mathbf{v}||^2_{\mathbf{A}} = \mathbf{v}^T\mathbf{A}\mathbf{v}$. The 122-element vectors $\mathbf{B}_m$ and $\mathbf{B}(\mathbf{x}, \mathbf{Q}(\mathbf{x}))$ represent the measured sensor activations, and sensor activations predicted by the forward model, respectively. Since $\mathbf{B}(\mathbf{x}, \mathbf{Q})$ is linear in $\mathbf{Q}$, that is, $\mathbf{B}(\mathbf{x}, \mathbf{Q}) = \mathbf{K}(\mathbf{x})\,\mathbf{Q}$, the dipole moment vector $\mathbf{Q}$ can be expressed by the least square method as:

$$\mathbf{Q}(\mathbf{x}) = (\mathbf{K}(\mathbf{x})^T\mathbf{K}(\mathbf{x}))^{-1}\mathbf{K}(\mathbf{x})^T\mathbf{B}_m.$$

where $\mathbf{K}(\mathbf{x})$ is the kernel of a spherical head model (Mosher et al., 1999) and $\Sigma$ is the noise covariance matrix, which is an identity matrix for spherical zero mean unit variance Gaussian noise. If the noise is known, the covariance matrix can be easily calculated. However the noise is generally unknown, so people often assume a spherical covariance matrix. Alternatively, one can measure the sensor activations before stimulation or long after stimulation and calculate the covariance matrix from those measurements. Since both of these techniques are popular, we simulate each. Figure 7 shows the mean localization error for 300 test sets of varying S/N, for both spherical and measured covariance matrices. Figure 7 shows that the closer the initial guess, the better the performance, under all tested covariance assumptions.

Figure 8 shows the mean localization error for all 3 300 activations with various S/N as the distance between the target and the initial guess is varied. Figure 8 tells us that with a good initial guess the measured covariance yields much better localization performance, but at the expense of performance when the initial guess is further from the target.

These experiments show that the localization of signals having high S/N ratios shows a larger dependence on the initial guess as the initial guess is moved away from the target!

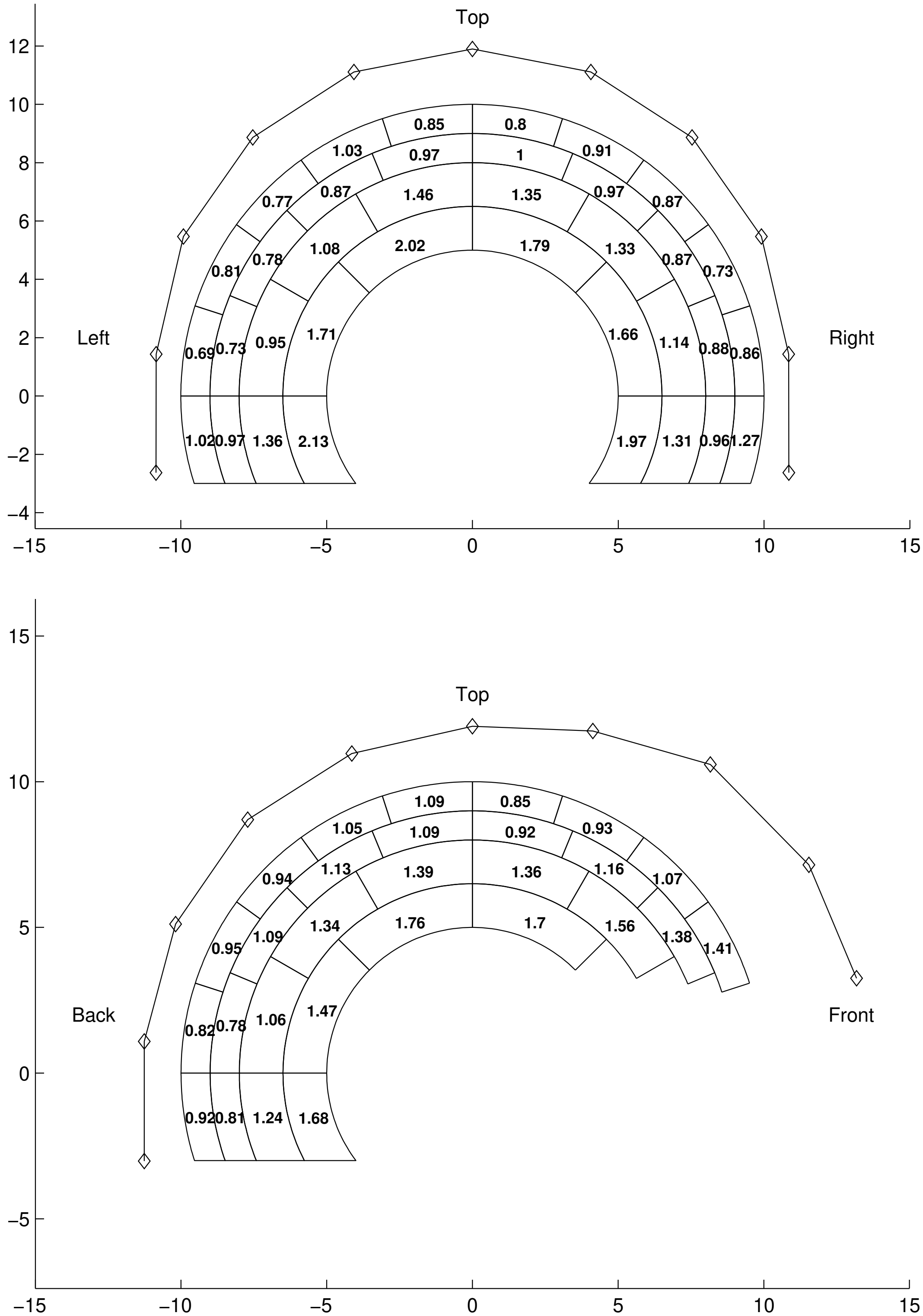


**Figure 6.** Mean localization errors of the MLP trained on the real brain noise dataset for dipole signals contaminated by real brain noise as a function of correct dipole location, binned into regions. All units are in cm. Regions not covered by the sensor surface in the saggital cross section were excluded. Top: Coronal cross section. Bottom: Saggital cross section.

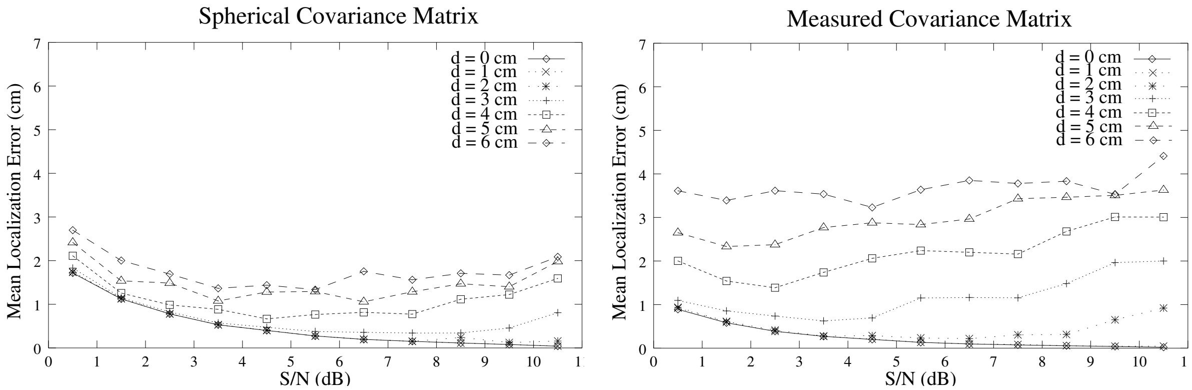


**Figure 7.** Mean localization error of 1-start-LM as a function of S/N at varying distances $d$ between the initial guess and the actual source dipole. Left: spherical covariance matrix. Right: measured covariance matrix.

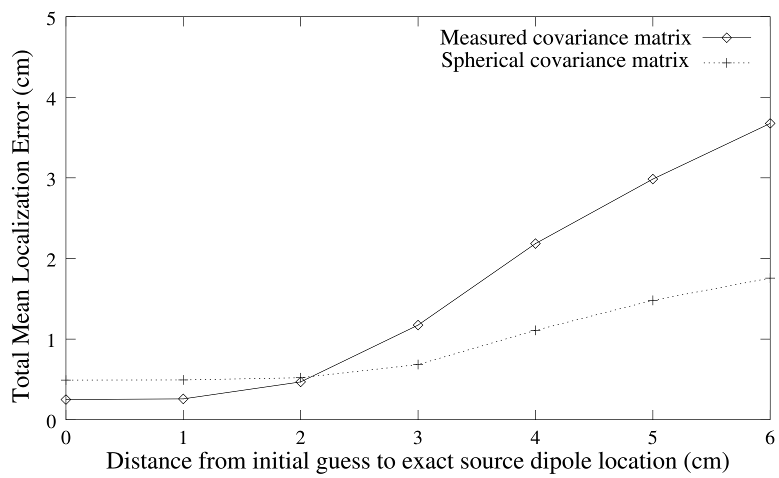


**Figure 8.** Localization error versus distance $d$ between LM using spherical covariance and LM using measured covariance. S/N ranges from 0–11 dB.

One can attempt to compensate for this effect by trying multiple restarts, yielding an $n$-start-LM algorithm. Localization performance of this method could level out quickly, with the randomly chosen initial guesses at about $n = 20$, as shown in Table 2.

These results motivated us to construct a hybrid system in which the MLP's output is used as the initial guess for LM. As we shall see in the next section, this MLP-start-LM performs very well indeed.

## 5. Comparative Performance: LM, MLP, MLP-start-LM

For the comparison between the MLP and the hybrid system, LM was used in three ways. First, LM was tuned for good performance using four re-starts at the fixed initial points $(0., -6.9282, 1.)$, $(-6., 3.4641, 1.)$, $(6., 3.4641, 1.)$, and $(0.01, 0.01, 6.1962)$, in units of cm. We call this system "fixed-4-start-LM." Second, LM was restarted with $n$ random initial guesses, for "random-$n$-start-LM." We checked how many re-starts of LM are needed to

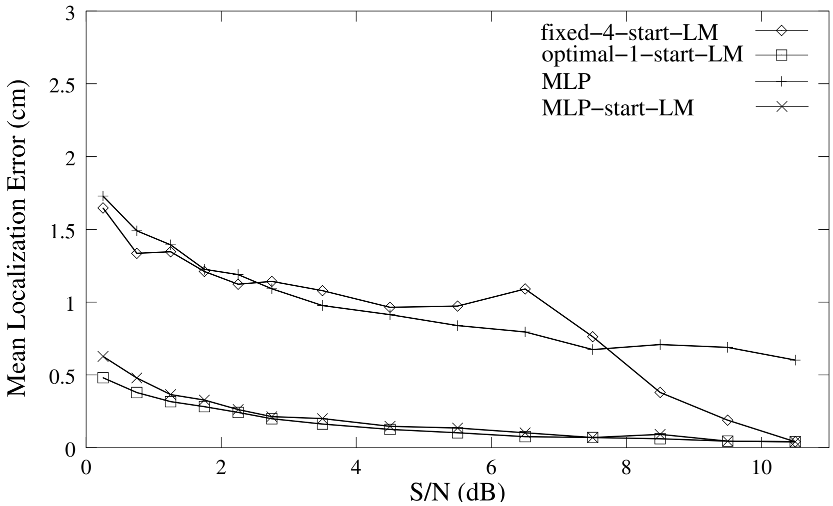


**Figure 9.** Mean localization error versus S/N. Each of fixed-4-start-LM, optimal-1-start-LM, MLP, and MLP-start-LM were tested on 4 500 dipole signals contaminated with real brain noise. The same MLP, trained with the real brain noise dataset, was used for both the MLP curve and the MLP-start-LM curve.

match the accuracy of the hybrid system. To match the hybrid system with a MLP trained with real brain noise, over 20 restarts of random-$n$-start-LM were required. Finally, we used the known exact dipole source location as the starting point, for "optimal-1-start-LM." For all three LM systems, the covariance matrix was calculated from real brain noise and used in the LM error measure.

Each of the MLP localizers from Section 3 was used as an LM initializer, for four variant MLP-start-LM localizers. The performance of three LM systems, all four MLP localization systems trained with various sorts of noise, and their hybrid systems, is shown in Table 2. With real brain noise, our MLP is the fastest: about fifteen hundred times faster than fixed-4-start-LM to match accuracies. Our hybrid MLP-start-LM is about four times more accurate and thirteen times faster than fixed-4-start-LM, while it is only slightly slower and less accurate than optimal-1-start-LM. Figure 9 shows localization performance as a function of S/N for fixed-4-start-LM, optimal-1-start-LM, the MLP trained with real brain noise, and its hybrid system. Optimal-1-start-LM shows the best localization performance across the whole range of S/N, and the hybrid system shows almost the same performance as the optimal-1-start-LM, while the trained MLP is more robust to noise than fixed-4-start-LM.

Table 2 shows that more than 20 random trials of LM were needed to match the accuracy of MLP-start-LM with real brain noise, so one might expect MLP-start-LM with real brain noise to be about twenty times faster than random-20-start-LM, and about four times faster than fixed-4-start-LM. However, MLP-start-LM is in fact about sixty times faster than random-20-start-LM, and about thirteen times faster than fixed-4-start-LM. The reason is that the initial guess of MLP-start-LM is closer to the point of convergence, so fewer iterations of LM are needed.

**Table 2.** Comparison of performance on real brain noise test set of Levenberg-Marquardt source localizers with three re-starts strategies; trained MLP; and hybrid system. Each number is an average over 4 500 localizations, so the error bars are negligible. The training used various sorts of noise (N=None, W=Uncorrelated White Gaussian, C=Correlated, B=Real Brain). Naturally performance is best when the training noise is drawn from the same distribution as the testing noise.

| algorithm | trained noise | computation time (ms) | localization error (cm) |
|---|---|---|---|
| fixed-4-start-LM | — | 449 | 1.16 |
| random-20-start-LM | — | 2 175 | 0.31 |
| optimal-1-start-LM | — | 22 | 0.23 |
| MLP | N | 0.3 | 2.70 |
| | W | 0.3 | 1.64 |
| | C | 0.3 | 2.06 |
| | B | 0.3 | 1.15 |
| MLP-start-LM | N | 53 | 0.84 |
| | W | 41 | 0.44 |
| | C | 49 | 0.67 |
| | B | 36 | 0.28 |

## 6. Conclusion

We have shown that the initial guess is critical for the Levenberg-Marquardt based localization method, and that LM performs much better with a good initial guess. The multilayer perceptron was shown to give good performance with reasonable accuracy across a range of mismatches between training and testing noise. The MLP's localization accuracy was comparable to fixed-4-start-LM's, at 1/1500-th the computational burden. This motivated us to construct a hybrid system, MLP-start-LM, which improves the localization accuracy beyond any other practical techniques available to us (by a factor of about four) while reducing the computational burden to less than a thirteenth that of fixed-4-start-LM. MLP-start-LM was comparable in accuracy to random-20-start-LM's, at 1/60-th the computation burden, which is about 3 times faster than might be naively expected. The hybrid system showed almost as good performance in accuracy and computation time as the hypothetical optimal-1-start-LM.

MLPs trained with various noisy datasets show better performance for dipole signals contaminated by real brain noise than those trained without noise. Even hybrid MLP-start-LMs initialized by MLPs trained with noise perform better than those initialized by the MLP trained without noise. Our conclusion is that MLPs should be trained with real noise, and can serve as real-time MEG dipole source localizers by themselves or as excellent automatic initial dipole guessors for iterative methods.

Our MLP was both trained and applied on data from a spherical volume conductor. One can expect a loss in accuracy if it is trained on a spherical but applied on a realistic volume conductor. However, since for MEG the spherical approximation is usually very good, we expect that the initial guess will remain very reasonable.

A number of extensions are planned in the immediate future: we will integrate a more sophisticated forward model already developed in our laboratory, we will experiment with different MLP structures and output representations in an effort to improve the accuracy of the MLP, and we will investigate the application of MLPs to the multi-dipole localization problem. We have yet to evaluate our system using real human subjects with dipoles in known locations.

## Acknowledgements

Supported in part by NSF CAREER award 97-02-311, the National Foundation for Functional Brain Imaging, and a gift from the NEC Research Institute.

## References

Abeyratne U R, Kinouchi Y, Oki H, Okada J, Shichijo F and Matsumoto K 1991 Artificial neural networks for source localization in the human brain *Brain Topography* **4** 3–21.

Hämäläinen M, Hari R, Ilmoniemi R J, Knuutila J and Lounasmaa O V 1993 Magnetoencephalography — theory, instrumentation, and applications to noninvasive studies of the working human brain *Reviews of Modern Physics* **65** 413–497.

Haykin S S 1994 *Neural Networks: A Comprehensive Foundation* Information and System Sciences Macmillan.

Hoey G V, Clercq J D, Vanrumste B, de Walle R V, Lemahieu I, D'Havé M and Boon P 2000 EEG dipole source localization using artificial neural networks *Phys. Med. Biol.* **45** 997–1011.

Hornik K, Stinchcombe M and White H 1989 Multilayer feedforward networks are universal approximators *Neural Networks* **2** 359–366.

Kamijo K, Kiyuna T, Takaki Y, Kenmochi A, Tanigawa T and Yamazaki T 2001 Integrated approach of an artificial neural network and numerical analysis to multiple equivalent current dipole source localization *Frontiers Med. Biol. Engng* **10** 285–301.

Kinouchi Y, Ohara G, Nagashino H, Soga T, Shichijo F and Matsumoto K 1996 Dipole source localization of MEG by BP neural networks *Brain Topography* **8** 317–321.

Kwon H C, Lee Y H, Jun S C, Kim J M, Park J C and Kuriki S 2000 Localization errors with 40-channel tangential fields *Biomag2000 Proc. 12th Int. Conf. on Biomagnetism* pp. 943–946.

Leahy R M, Mosher J C, Spencer M E, Huang M X and Lewine J D 1998 A study of dipole localization accuracy for MEG and EEG using a human skull phantom *Electroencephalography and clinical neurophysiology* **107**(2) 159–173.

LeCun Y, Kanter I and Solla S A 1991 Second order properties of error surfaces: Learning time and generalization *Advances in Neural Information Processing Systems 3* Morgan Kaufmann pp. 918–924.

Levenberg K 1944 A method for the solution of certain problems in least squares *Quart. Appl. Math* **2** 164–168.

Lütkenhöner B 1994 Magnetic field arising from current dipoles randomly distributed in a homogeneous spherical volume conductor *J. Appl. Phys.* **75** 7204–7210.

Marquardt D W 1963 An algorithm for least-squares estimation of nonlinear parameters *SIAM J. Appl. Math* **11** 431–441.

Mosher J C, Leahy R M and Lewis P S 1999 EEG and MEG: Forward solutions for inverse methods *IEEE Transactions on Biomedical Engineering* **46** 245–259.

Press W H, Flannery B P, Teukolsky S A and Verrerling W T 1988 *Numerical Recipes in C* Cambridge University Press.

Rumelhart D E, Hinton G E and Williams R J 1986 *Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1: Foundations* MIT Press.

Sarvas J 1987 Basic mathematical and electromaagnetic concepts of the biomagnetic inverse problem *Phys. Med. Biol.* **32** 11–22.

Sun M and Sclabassi R J 2000 The forward EEG solutions can be computed using artificial neural networks *IEEE Transactions on Biomedical Engineering* **47** 1044–1050.

Tang A C, Pearlmutter B A, Malaszenko N A, Phung D B and Reeb B C 2002 Localization of independent components of magnetoencephalography in cognitive tasks *Neural Computation* **14**(7). In press.

Tang A C, Pearlmutter B A, Zibulevsky M and Carter S A 2000 Blind separation of multichannel neuromagnetic responses *Neurocomputing* **32–33** 1115–1120.

Tang A C, Phung D, Pearlmutter B A and Christner R 2000 Localization of independent components from magnetoencephalography *International Workshop on Independent Component Analysis and Blind Signal Separation* Helsinki, Finland pp. 387–392.

Vigário R, Jousmäki V, Hämäläinen M, Hari R and Oja E 1998 Independent component analysis for identification of artifacts in magnetoencephalographic recordings *Advances in Neural Information Processing Systems 10* MIT Press pp. 229–235.

Yuasa M, Zhang Q, Nagashino H and Kinouchi Y 1998 EEG source localization for two dipoles by neural networks *Proc. 20th Ann. Int. Conf. of the IEEE Engineering in Medicine and Biology Society* pp. 2190–2192.

Zhang Q, Yuasa M, Nagashino H and Kinouchi Y 1998 Single dipole source localization from conventional EEG using BP neural networks *Proc. 20th Ann. Int. Conf. of the IEEE Engineering in Medicine and Biology Society* pp. 2163–2166.