MEG Source Localization using an MLP with a Distributed Output Representation

Sung Chan Jun* Barak A. Pearlmutter Guido Nolte

***Abstract*— We present a system that takes realistic magnetoencephalographic (MEG) signals and localizes a single dipole to reasonable accuracy in real time. At its heart is a multilayer perceptron (MLP) which takes the sensor measurements as inputs, uses one hidden layer, and generates as outputs the amplitudes of receptive fields holding a distributed representation of the dipole location. We trained this Soft-MLP on dipolar sources with real brain noise and converted the network's output into an explicit Cartesian coordinate representation of the dipole location using two different decoding strategies. The proposed Soft-MLPs are much more accurate than previous networks which output source locations in Cartesian coordinates. Hybrid Soft-MLP-start-LM systems, in which the Soft-MLP output initializes Levenberg-Marquardt, retained their accuracy of 0.28 cm with a decrease in computation time from 36 ms to 30 ms. We apply the Soft-MLP localizer to real MEG data separated by a blind source separation algorithm, and compare the Soft-MLP dipole locations to those of a conventional system.**

***Index Terms*— Magnetoencephalography, Source localization, Multilayer perceptron, Distributed representation**

I. Introduction

SOURCE localization using EEG and MEG identifies brain regions that emit detectable electromagnetic signals. The multilayer perceptron [1], a particular sort of universal approximator, has been recently used to build fast dipole localizers [2, 3, and references therein]. All this work used MLPs whose outputs represented source location or dipole moment vectors in *Cartesian* coordinates—a representation which might be expected to limit their performance and robustness.

We propose an MLP with a distributed representation[1] of the dipole location. Our Soft-MLP network, which uses that representation, localizes a dipole to reasonable accuracy in real time from MEG signals contaminated by considerable noise. Its output consists of the amplitudes of Gaussian receptive fields evenly distributed within a spherical head model, which taken together represent the dipole location. Like the Cartesian representation, this does not confine the dipole to a finite set of grid locations; but unlike the Cartesian representation, it is naturally tolerant to noise far from the region of interest.

II. Method

A. Data

The synthetic data used in our experiments consisted of corresponding pairs of dipole locations and sensor activations, as generated by a forward model. Given a dipole location and a set of sensor activations, the minimum error dipole moment can be calculated analytically [5]. Therefore, although the dipoles used in generating the data set have both location and moment, we discarded the moment in all the experiments below.[2]

We made two datasets, one for training and the other for testing. Dipoles were drawn uniformly from truncated spherical regions [3, Figure 1]. Their moments were drawn uniformly from vectors of strength ≤ 100 nA·m. The corresponding sensor activations were calculated by adding the results of a forward model and a noise model. To make sure the network does not inappropriately project external sources into the brain,[3] and allow the network to interpolate rather than extrapolate thus improving performance, the training set used dipoles from the larger region, while to better approximate field conditions the test set contained only dipoles from the smaller inner region. We used the sensor geometry of a 4D Neuroimaging Neuromag-122 whole-head MEG system[4] [7], and an analytic forward model of quasistatic electromagnetic propagation in a spherical head [3, Section 2.1].

In order to properly 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 collected real brain noise from unaveraged MEG recordings (task involving abrupt visual stimulation and subsequent brief motor output and audio feedback, two right-handed myopic middle-aged female subjects, analog bandpass filter 0.03–100 Hz) during periods far from the stimulus or response. This noise had a square root mean square (RMS) sensor reading of $P^{\mathrm{n}} \approx$ 50–100 fT/cm. We measured the SNR of a dataset using the ratios of the powers in the signal and noise, SNR (in dB) $= 20 \log_{10} P^{\mathrm{s}}/P^{\mathrm{n}}$, where P^{s} is the RMS sensor reading from the dipole.

B. Soft-MLP structure

The Soft-MLP charged with approximating the inverse mapping had an input layer of 122 units, one for each sensor; one hidden layer with N units; and an output layer of $M = 193$ units representing the amplitudes of three-dimensional Gaussian receptive fields in the training region of the head model. The target output representation of a dipole at location $\mathbf{x}$ was the M-dimensional vector $\mathbf{G}(\mathbf{x})$ defined by $G_i(\mathbf{x}) = \exp -||\mathbf{x} - \mathbf{x}_i||^2/2\sigma^2$ where $\mathbf{x}_i$ is the center of Gaussian receptive field i and σ is the length scale of the receptive fields. The receptive field centers $\mathbf{x}_i$ were *evenly distributed* with a spacing of 3 cm, and we set $\sigma = 1.8$ cm. These parameters were determined empirically. With these, $M = 193$ receptive fields served to cover the training region.

The output units had linear activation functions[5], while to accelerate multitask training [6], we found no decrease in localization error. Typically an accurate estimate of the location is much more important than of the moment direction or strength. We also found that networks trained without the moment were more robust.

Supported by NSF CAREER award 97-02-311, the MIND Institute, and the NEC Research Institute. *Asterisk indicates corresponding author.*

*Sung Chan Jun, Barak A. Pearlmutter and Guido Nolte are with the Department of Computer Science, University of New Mexico, Albuquerque, NM 87131, USA (Email: {junsc,bap,nolte}@cs.unm.edu).

[1] The term *distributed representation* is standard in neural networks [4].

[2] We experimented with a dataset containing as targets both the location and moment of each dipole, and despite the increased generalization expected for

[3] This can easily occur in practice, for instance when the head position is incorrectly measured. This is a condition we would like our system to note, rather than silently projecting external dipoles into the region it believes to be occupied by the brain.

[4] This MEG system has 61 pairs of first-order planar gradiometers.

[5] In artificial neural networks, the *activation function* computes the output value of an artificial neuron based on the weighted sum of its inputs. The output value may be continuous or discrete, and Heavyside, linear, sigmoid, and hyperbolic tangent activation functions are widely used.

ate training the hidden units had hyperbolic tangent activation functions [8]. Adjacent layers were fully connected, and there were no cut-through connections. Input data were usually preprocessed to improve the performance [2], and here the 122 MEG sensor activation inputs were scaled to an RMS value of 0.5. The network weights, $\mathbf{w}$, were initialized with uniformly distributed random values between ± 0.1. Backpropagation was used to calculate the error gradient [1], which in turn was used for online stochastic gradient decent optimization with an empirically chosen descent rate η, as in $\Delta \mathbf{w} = -\eta \nabla_{\mathbf{w}} E$.

In determining a reasonable MLP structure, practical considerations constrained our experiments to networks with no more than 160 hidden units. We ran experiments with 20, 40, 60, 80, 120, and 160 units in the hidden layer. Each MLP was trained with noise-free training datasets of size 500, 1 000, 2 000, 4 000, 8 000, 16 000, and 32 000. For each MLP the best generalization error[6] in 500 epochs of training[7] was measured, using a noise-free test set of 5 000 patterns. For each MLP size and training dataset five runs were performed, and the generalization errors averaged.

The computation time for localization increases linearly with the number of hidden units, and the training time increases about linearly with the size of training dataset and the size of the hidden layer. When the training set is small, generalization error is high. Increasing the computation, *i.e.* increasing the size of the training set or the number of units in a hidden layer, tends to reduce the generalization error, nearing an asymptote at about 8 000 exemplars and 80 hidden units. For this reason we chose to use 80 hidden units.

C. Decoding strategies

For practical use, and to measure performance, the M-dimensional distributed representation of the dipole location must be converted to Cartesian coordinates. We experimented with two strategies for decoding the output vector, $\mathbf{a} = (a_1, a_2, \ldots, a_M)$, under the assumption that $a_i \approx G_i(\mathbf{x})$.

Strategy 1:

- Find the index of the receptive field with the maximum amplitude, $i^* = \text{argmax}_i a_i$.
- Linearly interpolate between the centers of the receptive fields in a ball B_{i^*} with center $\mathbf{x}_{i^*}$ and radius 6 cm (twice the inter-center distance)[8] using the activation values as weights, $\hat{\mathbf{x}} = \sum_{\mathbf{x}_i \in B_{i^*}} a_i \mathbf{x}_i / \sum_{\mathbf{x}_i \in B_{i^*}} a_i$.

Strategy 2:

- For each of the M receptive field centers place a ball B_i with center $\mathbf{x}_i$ and radius 6 cm (twice the inter-center distance), and calculate $c_i = a_i + \sum_{j \neq i, \mathbf{x}_j \in B_i} a_j / ||\mathbf{x}_i - \mathbf{x}_j||$.
- Find $i^* = \text{argmax}_i c_i$.
- Apply the linear interpolation of Strategy 1 step 2.

III. Localization Results

A. Comparison of Soft-MLP and Cartesian-MLP

The training dataset contained 20 000 exemplars, contaminated with real brain noise, and another dataset, of 4 500 MEG signal patterns contaminated by real brain noise, was constructed for testing.[9] The distribution of SNRs for patterns in the testing dataset are shown in Table I. We trained for up to 500 epochs, which took about twelve hours on an 800 MHz AMD Athlon for each training dataset.

TABLE I
Distribution of SNR for the 4 500 testing patterns.

SNR (dB)	# of Patterns	Frequency (%)
0–2	892	19.82
2–4	806	17.91
4–6	824	18.31
6–8	627	13.93
8–10	538	11.96
10–12	381	8.47
12–14	229	5.09
> 14	203	4.51

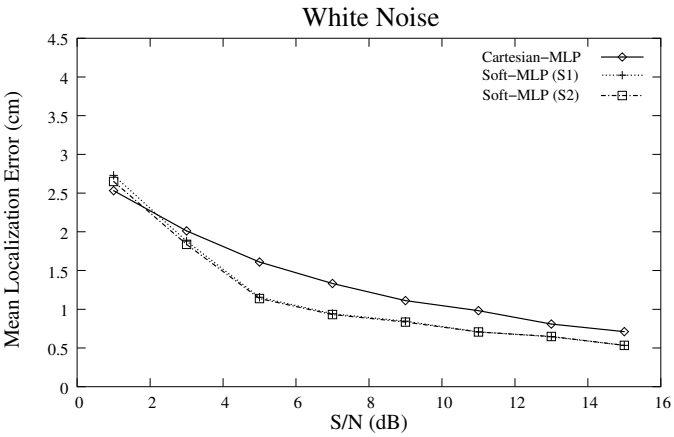

Fig. 1. Mean localization error *vs.* SNR for the Cartesian-MLP, the Soft-MLP using two sorts of decoding strategies, and their hybrid methods. Decoding strategy 1 and strategy 2 are denoted by S1 and S2, respectively.

After each Soft-MLP was trained, its performance in RMS linear accuracy was measured by converting the outputs to Cartesian coordinates. Using each of the two decoding strategies led to two different systems. The Soft-MLP performance, along with that of a Cartesian-MLP network,[10] is shown as a function of input SNR in Figure 1. As a whole, the Soft-MLP is more accurate than the Cartesian-MLP. The Soft-MLP using Strategy 2 shows a slight performance advantage over Strategy 1. Each of the Soft-MLP localizers was used as an LM initializer, giving two variant Soft-MLP-start-LM hybrids. Their performance is shown as a function of SNR and compared with the hybrid method using the Cartesian-MLP in Figure 1. The MLP-start-LMs of the toft-MLPs show better localization accuracy at high SNRs than the MLP-start-LM of the Cartesian-MLP, while they have degraded accuracy at low SNRs. These results also held for networks trained with other sorts of noise (not shown).

A grand summary, averaged across various SNR conditions, is shown in Table II. In comparing the Soft-MLP (S2) and Cartesian-MLP localizers, both trained with real brain noise, one sees that localization error improved from 1.15 cm to 0.85 cm, while computation time increased from 0.3 ms to 1.0 ms. With an increased expense in

[6] Because our training sets were large, the performance of the network on the training dataset and on a new testing dataset would be about the same, were the testing dataset not taken from a smaller region. The test dataset is used to verify the system performance, and *generalization error* means the average error obtained on a test dataset.

[7] In one epoch each exemplar in the training dataset is presented once.

[8] When a bigger radius is used, outliers aren't filtered out and computation is more costly. When a smaller radius is used, significant values might be thrown away. Empirically, twice the inter-center distance balanced these consideration.

[9] The number of exemplars was constrained in part by the availability of real brain noise data in which we were confident.

[10] The structure of Cartesian-MLP was empirically optimized by trading off computation and accuracy [3].

TABLE II
PERFORMANCE OF CARTESIAN-MLPS, SOFT-MLPS, AND MLP-START-LM HYBRIDS. EACH NUMBER IS AN AVERAGE OVER 4 500 LOCALIZATIONS. SOFT-MLPS WERE TESTED USING TWO DECODING STRATEGIES (S1/S2).

Algorithm	Time (ms)	Accuracy (cm)
Cartesian-MLP	0.3	1.15
Soft-MLP (S1)	0.7	0.87
Soft-MLP (S2)	1.0	0.85
MLP-start-LM (Cartesian)	36	0.28
MLP-start-LM (Soft, S1)	31	0.30
MLP-start-LM (Soft, S2)	30	0.28

time, the distributed output representation yielded much more accurate (assuming a spherical uncertainty, the zone in which the dipole is likely located is decreased from 1.5 cm^3 to 0.6 cm^3, a factor of 2.5) localizations. The MLP-start-LM method using Soft-MLP (S2) has the same localization error as Cartesian-MLP, 0.28 cm. However, it is slightly *faster!* This surprising reduction in total computation time is due to the Soft-MLP generating a better initial guess, resulting in fewer iterations of LM.

B. Localization for actual BSS-separate MEG components

We applied the Soft-MLP (S2) trained with real brain noise to localize dipolar sources from actual BSS-separated MEG signal components. The xfit program (standard commercial software bundled with the 4D Neuroimaging Neuromag-122 MEG system) is compared with the methods developed here. We chose eight of the actual BSS-separated MEG signal components from which xfit localized a single dipole source well, and which met other criteria for correct localization laid out in [9]. (Continuous MEG data were collected, sampled at 300 Hz, band-pass filtered at 0.03–100 Hz, separated using SOBI, and scanned for neuronal sources of interest. See [9, 10] for full details.)

Figure 2 shows the localized dipoles from three viewpoints: axial (x-y plane), coronal (x-z plane), and saggital (y-z plane). The MLP-estimated locations are about 1.18 cm on average from those of xfit. The trained Soft-MLP is applicable to actual MEG signals, and can be a good initial guessor for iterative methods with clear advantages in speed and in the lack of required human interaction.

C. Comparison with the global search algorithm

The global search algorithm uses storage to reduce computation in dipole localization [11]. Briefly: a number of grid points are selected in the head model, and the field pattern at the sensors resulting from orthogonally oriented dipoles at each location are precomputed. This information allows the orientation and strength of a dipole located at a particular grid point which best fits a vector of measurements to be efficiently calculated. When a measured signal is to be localized, the goodness of fit of the best dipole at each grid point is calculated, and the location of the grid point with the best GOF is used to initialize a gradient-based optimization routine.

This table-based algorithm is surprisingly efficient at localizing dipolar sources. For example, it has been used to localize a dipole at each time point in a large (100 000 sample) MEG dataset [11]. The primary weakness of the global search algorithm is that the gridding must be fine enough for the problem at hand. In particular, the spacing of precomputed points must be well above the Nyquist limit of the highest spatial frequency in the error surface, or the correct optimum can be skipped over. Therefore, table size, and therefore the time required for a localization, will increase with increasing complexity of the error surface. The error surface might become more complex under two circumstances: (a) with a more realistic head model [12], and (b) with a complex MEG helmet, for instance a helmet with superconducting magnetic reflectors [13].

In contrast, the Soft-MLP is robust to a complex error surface. Even with a realistic head model or a highly complex MEG helmet, Soft-MLP is applicable without modification. Under such circumstances, we expect the Soft-MLP to do a single localization much more quickly than the global search algorithm, even though it may take longer to train the MLP than to precompute the global search algorithm's tables.

Another advantage of the Soft-MLP is that it can be trained using an arbitrary noise model, characterized only by a set of samples, such as actual measured brain noise. Because the global search algorithm relies on a least-squares fit to determine the GOF at each grid point, its noise model must be Gaussian.

IV. SUMMARY

We propose the use of distributed representations to encode dipole locations in the output of MLP-based dipole localizers. Experiments showed that such a network was fast and robust, and was a better dipole source localizer (0.85 cm *vs.* 1.15 cm) at slightly greater computational expense (1.0 ms *vs.* 0.3 ms) than a comparably tuned system using a Cartesian representation. The hybrid MLP-start-LM method using the new MLP showed the same accuracy as previous systems (0.28 cm) but computation time was reduced from 36 ms to 30 ms. Furthermore, the Soft-MLP was successfully applied to actual MEG data.

A Cartesian output representation cannot encode the location of more than a single dipole. Our use of a distributed output representation was in part motivated by the hope that its greater representational capabilities might allow Soft-MLP networks to be used for multiple dipole localization. The improvements in accuracy for a single dipole were an unexpected benefit, but we will continue our efforts to apply the Soft-MLP architecture to the multiple dipole case.

REFERENCES

[1] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back–propagating errors. *Nature*, 323:533–536, 1986.

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

[3] Sung Chan Jun, Barak A. Pearlmutter, and Guido Nolte. Fast accurate MEG source localization using a multilayer perceptron trained with real brain noise. *Physics in Medicine and Biology*, 47(14):2547–2560, July 21 2002.

[4] Geoffrey E. Hinton, James L. McClelland, and David E. Rumelhart. Distributed representations. In D. E. Rumelhart, J. L. McClelland, and the PDP research group., editors, *Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1: Foundations*. MIT Press, 1986.

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

[6] Richard Caruana. Multitask learning. *Machine Learning*, 28(1): 41–75, July 1997.

[7] A. I. Ahonen, M. S. Hämäläinen, J. E. T. Knuutila, M. J. Kajola, P. P. Laine, O. V. Lounasmaa, L. T. Parkkonen, J. T. Simola, and C. D. Tesche. 122-channel SQUID instrument for investigating

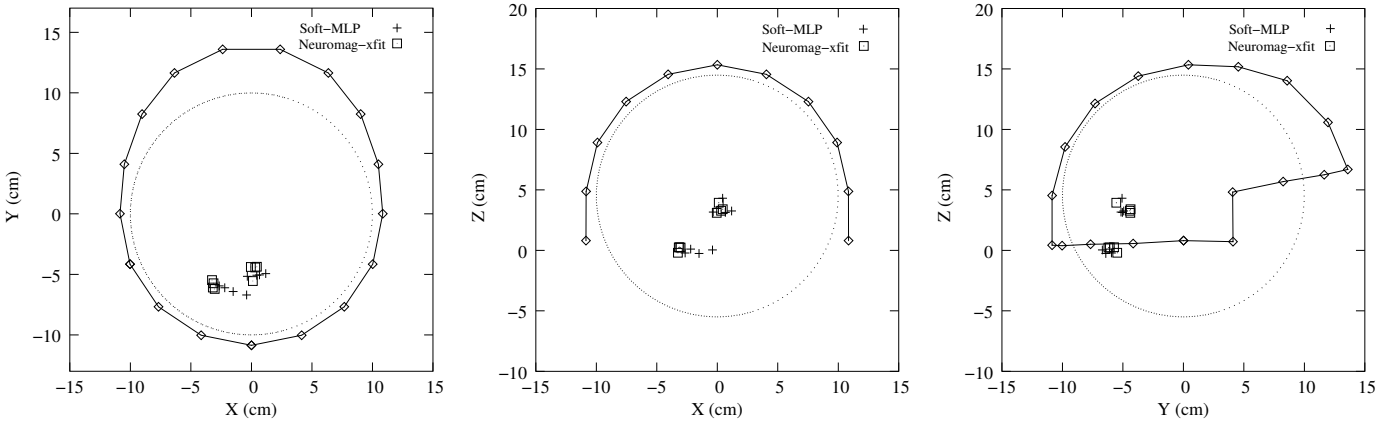

Fig. 2. Dipole source localization results, using real data with sources separated using SOBI. Locations found by standard Neuromag xfit software *vs.* the Soft-MLP (S2) for eight actual BSS-separated visual MEG components. Left: Axial view. Center: Coronal view. Right: Saggital view. The outer surface denotes the sensor surface, and diamonds on the surface denote sensors. The inner surface denotes the spherical head model.

the magnetic signals from the human brain. *Physica Scripta*, T49:198–205, 1993.

[8] Yann LeCun, Ido Kanter, and Sara A. Solla. Second order properties of error surfaces: Learning time and generalization. In *Advances in Neural Information Processing Systems 3*, pages 918–924. Morgan Kaufmann, 1991.

[9] Akaysha C. Tang, Barak A. Pearlmutter, Natalie A. Malaszenko, Dan B. Phung, and Bethany C. Reeb. Independent components of magnetoencephalography: Localization. *Neural Computation*, 14(8):1827–1858, August 2002.

[10] Akaysha C. Tang, Barak A. Pearlmutter, Michael Zibulevsky, Tim A. Hely, and Michael P. Weisend. An MEG study of response latency and variability in the human visual system during a visual-motor integration task. In *Advances in Neural Information Processing Systems 12*, pages 185–191. MIT Press, 2000. URL http://nips.djvuzone.org/djvu/nips12/0185.djvu.

[11] J. C. de Munck, A. de Jongh, and B. W. van Dijk. The localization of spontaneous brain activity: An efficient way to analyze large data sets. *IEEE Transactions on Biomedical Engineering*, 48:1221–1228, 2001.

[12] R. Van Uitert, D. Weinstein, C. Johnson, and L. Zhukov. Finite element EEG and MEG simulations for realistic head models: quadratic vs. linear approximations. *Journal of Biomedizinische Technik*, 46:32–34, 2001.

[13] R. H. Kraus, Jr., P. L. Volegov, K. Maharajh, M. A. Espy, A. N. Matlashov, and E. R. Flynn. Performance of a novel SQUID-based superconducting imaging-surface magnetoencephalography system. *Physica C*, 368(1-4):18–23, March 2002.