ENSEMBLE DE-NOISING OF SPATIO-TEMPORAL EEG AND MEG DATA

Stephan Weiß[1,2], *Richard M. Leahy*[2], *John C. Mosher*[3], *and Robert W. Stewart*[1]

[1] Signal Processing Division, Univ. of Strathclyde, Glasgow, UK
[2] Signal and Image Processing Institute, Univ. of Southern California, Los Angeles, CA
[3] Biophysics Group, Los Alamos National Laboratory, NM

ABSTRACT

We present a de-noising method for spatio-temporal EEG/ MEG data that incorporates prestimulus and spatial information to estimate the noise and signal energies and select a robust de-noising threshold. Improved performance over standard de-noising is demonstrated.

1. INTRODUCTION

The electro- and magnetoencephalogram (EEG/ MEG) are recordings of the scalp potential and magnetic field outside the head reflecting neural activity in the brain. The transient neural current sources associated with event related EEG/ MEG are generally assumed focal in nature and can be approximated using equivalent current dipoles. Using measurements from a sensor array, the location, orientation, and time series of a number of dipoles can be determined by solving an inverse problem [7]. However, even after stimulus-locked averaging from multiple trials, the data often still possesses a very low SNR due to background brain activity and instrumental and environmental noise. This low SNR results in reduced accuracy of the estimated dipole parameters [8].

Our goal is to improve the SNR of recorded EEG/ MEG data by wavelet de-noising which is a method for recovering an unknown transient signal from broadband noise [5]. By applying a suitable unitary transform to the signal, the noise will remain spread across the transform space while the signal can be parameterized by a few transform coefficients that will stand out from the noise. The suitability of the discrete wavelet transform (DWT) for parameterizing transient evoked responses within the EEG has been demonstrated in, for example, [1, 11]. By applying an inverse transform after appropriately thresholding the transform coefficients, noise reduction can be achieved. In the following, we introduce a form of de-noising that has been adapted to our problem, and present some simulation results to demonstrate the benefit of this approach to noise reduction.

In our notation, we will use normal letters for scalar values, and bold face lower and upper case letters for vector and matrix quantities, respectively.

2. SIGNAL MODEL

We assume that r dipolar neural sources are activated in the brain in response to a particular stimulus. The signal at an M dimensional sensor array is formed by the superposition of the fields at each of the i dipoles with position $\mathbf{r}_i$, orientation $\mathbf{q}_i$ and time series $\mathbf{s}_i$, with instrumental and environmental noise to yield an M-by-N spatio-temporal data matrix

$$\mathbf{F} = \sum_{i=0}^{r-1} \mathbf{G}(\mathbf{r}_i)\mathbf{q}_i\mathbf{s}_i^T + \mathbf{N} = \mathbf{X} + \mathbf{N}, \qquad (1)$$

with
$\mathbf{G}(\mathbf{r}_i) \in \mathbb{R}^{M\times 3}$ gain matrix,
$\mathbf{q}_i \in \mathbb{R}^3$ dipole orientation,
$\mathbf{s}_i \in \mathbb{R}^N$ dipole time series, and
$\mathbf{N} \in \mathbb{R}^{M\times N}$ additive noise.

The matrix $\mathbf{G}(\mathbf{r}_i)$ contains the gain factors from the ith dipole at location $\mathbf{r}_i$ to the M sensors.

If we assume the noise in $\mathbf{N}$ to be zero-mean Gaussian and uncorrelated with the source transients, then we can make the following approximation:

$$\|\mathbf{F}\|_F^2 \approx \|\mathbf{X}\|_F^2 + \|\mathbf{N}\|_F^2 \quad . \qquad (2)$$

The noise power can be estimated from prestimulus data, corresponding to a period of $\tilde{N}$

time slices before any event-related signals are produced, yielding

$$\|\mathbf{N}\|_F^2 \approx \frac{N}{\tilde{N}} \|\tilde{\mathbf{N}}\|_F^2 \quad . \tag{3}$$

A signal-to-noise ratio (SNR) of EEG/ MEG data can then be calculated as

$$\mathrm{SNR} = \frac{\|\mathbf{X}\|_F^2}{\|\mathbf{N}\|_F^2} \tag{4}$$

using the Frobenius norm $\|\cdot\|_F$.

3. ENSEMBLE DE-NOISING

De-noising as originally proposed by Donoho and Johnstone [5] is applicable to 1-dimensional signals corrupted by white noise. The signal is transformed, the transform coefficients thresholded according to some (heuristic) criterion, and an inverse transform used to obtain a noise reduction. Usually a discrete wavelet transform (DWT) is employed as its ability to yield local representation in both time and frequency domain is advantageous for the analysis of transient signals. Here, we represent a DWT by a unitary matrix $\mathbf{T}$, such that $\mathbf{y} = \mathbf{T}\mathbf{x}$ is the DWT of $\mathbf{x}$ [9]. Applying the transform to the temporal dimension of the data matrix $\mathbf{F}$, we can express the de-noising procedure as

$$\tilde{\mathbf{F}} = \Theta\big(\mathbf{F}\mathbf{T}^T\big)\,\mathbf{T}, \tag{5}$$

where $\Theta(\cdot)$ is the threshold operator and $(\cdot)^T$ denotes transpose.

For spatio-temporal EEG/ MEG data, the signal of interest in each channel is a linear combination of the same source transients. We can improve the performance of the de-noising procedure by making use of this property, i.e. since we expect to see similar characteristics in the transform coefficients across the sensor array [10] we can use a mask common to the whole array rather than de-noising each measurement time series separately, ie.

$$\tilde{\mathbf{F}} = \mathbf{F}\ \underbrace{\mathbf{T}^T\mathbf{M}\mathbf{T}}_{\text{TF-filter}} \tag{6}$$

with $\mathbf{M} = \mathrm{diag}\{\mu_n\} \in \mathbb{R}^{N\times N}$. For hard thresholding, the elements of $\mathbf{M}$ form a binary mask

$$\mu_n = \begin{cases} 1 & : \ \bar{t_f^2}(n) \geq \theta \\ 0 & : \ \bar{t_f^2}(n) < \theta \end{cases} \tag{7}$$

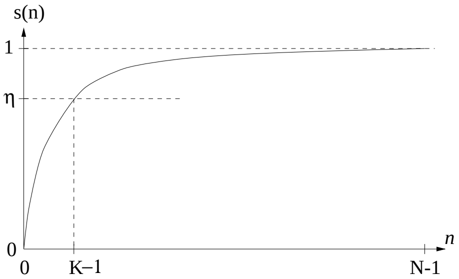

Fig. 1: Threshold selection for ensemble de-noising.

depending on the squared transform coefficients averaged over the spatial dimension

$$\bar{t_f^2}(n) = \frac{1}{\|\mathbf{F}\mathbf{T}^T\|_F^2} \sum_{m=0}^{M-1} |t_f(m,n)|^2, \tag{8}$$

with a normalization $\|\mathbf{F}\mathbf{T}^T\|_F^2$ to ensure $\|\bar{\mathbf{t}_f^2}\|_1 = 1$. We want to select the threshold θ by picking the K largest coefficients from $\bar{\mathbf{t}_f^2}$ such that the de-noised signal only retains as much energy as that estimated for the signal using (4). Next the elements of $\bar{\mathbf{t}_f^2}$ are re-ordered by a permutation matrix $\mathbf{P}$ such that the resulting vector

$$\tilde{\mathbf{t}_f^2} = \mathbf{P}\bar{\mathbf{t}_f^2} \tag{9}$$

has its elements in descending order, i.e. $\tilde{t_f^2}(0) \geq \tilde{t_f^2}(1) \geq \cdots \geq \tilde{t_f^2}(N-1)$. Then the cumulative sum on $\tilde{t_f^2}(n)$

$$s(n) = \sum_{i=0}^{n} \tilde{t_f^2}(i), \tag{10}$$

represents the normalized energy in the n largest coefficients. We determine the number of retained coefficients, K, by comparing $s(n)$ to the estimated SNR, η, between the energies of $\mathbf{X}$ and $\mathbf{F}$ as illustrated in Fig. 1. The equivalence of the energy of the spatio-temporal data and that of the transform coefficients is a direct result of the norm preserving property of the unitary transform $\mathbf{T}$. The ratio η can be calculated from the EEG/ MEG data using assumptions (2) and (3):

$$\eta := \frac{\|\mathbf{X}\|_F^2}{\|\mathbf{F}\|_F^2} \approx \frac{\|\mathbf{F}\|_F^2 - \frac{N}{\tilde{N}}\|\tilde{\mathbf{N}}\|_F^2}{\|\mathbf{F}\|_F^2}. \tag{11}$$

We then search for the minimum $K \in \mathbb{N}$ that fulfills $s(K-1) \geq \eta$, and obtain the threshold $\theta = \tilde{t_f^2}(K-1)$.

Instead of checking condition (7) for every $\bar{t}_f^2(n)$, it is possible to directly calculate the diagonal elements of $\mathbf{M}$, arranged in a vector $\boldsymbol{\mu}$, by recovering the indices of the K largest transform coefficients by

$$\boldsymbol{\mu} = \mathbf{P}^T \tilde{\boldsymbol{\mu}}, \qquad (12)$$

after initializing $\tilde{\mu}_n$ to

$$\tilde{\mu}_n = \begin{cases} 1 & : \quad n \le K \\ 0 & : \quad n > K \end{cases} \qquad (13)$$

The inverse transform is then applied to the K largest coefficients, computed as described above, to reconstruct the de-noised signal. We refer to this method as ensemble de-noising, since the spatial dimension of the data presents us with ensemble probes of the noise process.

Implicit in this procedure is the assumption that the signals of interest, ie. the dipole time series, can be represented by a small subset of the K basis functions of the transform. The quality of the parameterization determines the ability to recover a minimally distorted de-noised signal. If we further assume iid Gaussian noise, a parameterization by K coefficients yields a noise reduction of approx. K/N. However, even if the analysis function closely matches the source time series, problems arise due to phase sensitivity through the cyclo-stationarity of the DWT associated with the subsampling process. In [3], this is addressed by a translation-invariant approach (TI de-noising), where circularly shifted versions of the data matrix are ensemble de-noised, back-shifted, and averaged, thus effectively avoiding subsampling and its problems at the expense of increased computational complexity.

4. RESULTS AND DISCUSSION

We compare the previously described de-noising methods to a low rank approximation of the data matrix via a truncation of the SVD expansion [6], which is implicitly performed in the source localization algorithm in [7]. For spatially and temporally white noise, the SVD method yields a reduction of noise variance from $\mathbf{F}$ to $\tilde{\mathbf{F}}$ by a factor of r/R, where $R = \min(M, N)$ and r is the rank of the noise-free matrix $\mathbf{X}$ and equals the number of temporally independent dipolar sources. In comparison, de-noising gives a noise variance reduction of K/N as described above.

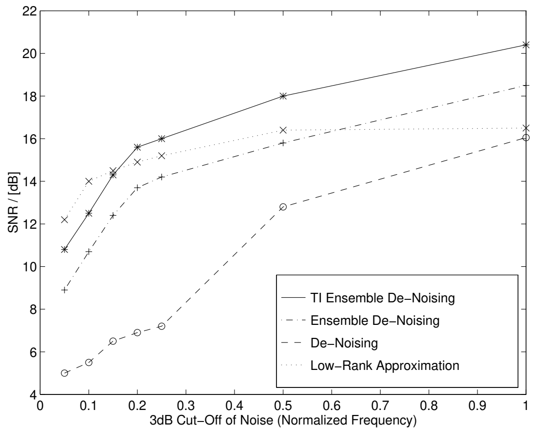

Fig. 2: Comparison of different de-noising methods applied to data from a 73-sensor array. SNR is 5dB, with temporally coloured noise of different bandwidths produced using a 5th order Butterworth filter. The signal of interest arises from three dipolar sources activated by Hermite-type functions. All de-noising methods use a Symmlet-8 wavelet for the transform.

Fig. 2 shows the averaged results over 25 trials with different noise power spectra and realistic source data for different de-noising techniques and the low rank SVD approximation. The colouring is achieved by lowpass filtering, with the cut-off frequency indicated on the abscissa. For standard de-noising of each sensor time series, we heuristically achieved best results using a visu-shrink soft threshold [2]. The SNR yielded by ensemble de-noising is consistently higher due to exploitation of pre-stimulus and spatial information. Apparently, a low rank approximation — ideal in the sense of knowing the exact number of independent dipoles — performs more robustly for strongly coloured noise. However, note that further improvements are achieved by applying translation-invariant (TI) ensemble de-noising.

Simulation results for different SNRs of $\mathbf{F}$ are presented in Fig. 3. Here $\mathbf{X}$ stems from a single dipolar source activated by the analysis wavelet, thus giving the best possible result of the presented method. The steep drop in performance of ensemble de-noising for low SNR is due to a high probability for any coefficient to pass the threshold θ. A second important observation is that ensemble de-noising followed by a low-rank approximation increases noise reduction over ei-

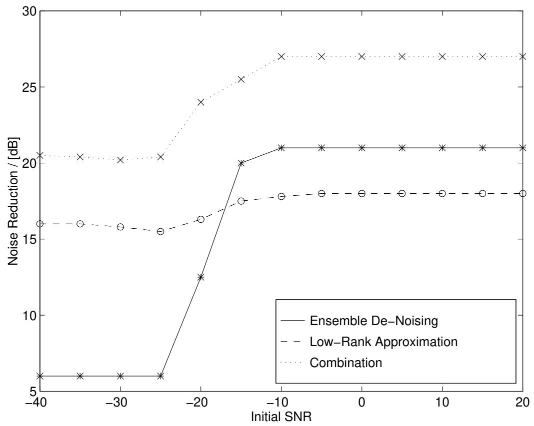

Fig. 3: Comparison of combination of ensemble de-noising with a low-rank approximation for different SNRs.

ther method separately. Since subspace-based dipole source localization algorithms inherently perform a low-rank approximation [7], ensemble de-noising is expected to give additional benefit.

For the situation in Fig. 3, Tab. 1 contains results of source parameter estimation [7] for different SNRs with and without ensemble de-noising (EDN), with small improvements in estimated location (in cm) and more significant improvements in time series (with $\|\mathbf{s}\|_2 = 1$) when using denoising. We found that, compared to low-rank approximation, denoising is very good at removing noise from intervals where no source is active but does not perform well when if the analysis wavelet and source activation function are not well matched.

Since good parameterization is crucial, in our current research we are looking into methods for further improvement of ensemble de-noising by adapting the transform to the analyzed data. Coifman *et al.* [4] have introduced a best basis selection method for the transform, such that the signal energy is contained in as few coefficients as possible. With a similar criterion, a library of different wavelets can be searched to find a basis function that most closely matches the signal's features and subsequently leads to a low entropy in the transform domain. There also is the possibility of using soft thresholding in (6), with an appropriate threshold function designed to preserve the estimated energy of the signal of interest.

SNR	EDN	$\|\overline{\Delta\mathbf{r}_n}\|_2$	$\|\overline{\Delta\mathbf{s}_n}\|_2$
5dB	no	0.0423	0.0154
	yes	0.0410	0.0109
0dB	no	0.1012	0.0511
	yes	0.0982	0.0361
-5dB	no	0.2944	0.1366
	yes	0.2442	0.0879

Tab. 1: Averaged deviation of location (in cm) and time series error for source parameter estimation according to [7] for a single dipole.

5. REFERENCES

[1] E. Bartnik, K. Blinowska, and P. Durka. "Single Evoked Potential Reconstruction by Means of Wavelet Transform". *Biol. Cybernetics*, Vol.67:pp.175–181, 1992.

[2] J. Buckheit, S. Chen, D. Donoho, I. Johnstone, and J. Scargle. *WaveLab 7.01*. Stanford University, 1996.

[3] R. R. Coifman and D. L. Donoho. "Translation-Invariant De-Noising". In A. Antoniades, editor, *Wavelets and Statistics*. Springer Verlag, to appear.

[4] R. R. Coifman and M. V. Wickerhauser. "Entropy-Based Algorithms for Best Basis Selection". *IEEE Trans. Inf. Theory*, Vol.38(No.2):pp.713–718, March 1992.

[5] D. L. Donoho and I. M. Johnstone. "Ideal Spatial Adaptation Via Wavelet Shrinkage". *Biometrika*, Vol.81:pp.425–455, 1994.

[6] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 3rd ed., 1996.

[7] J. C. Mosher, P. S. Lewis, and R. M. Leahy. "Multiple Dipole Modeling and Localization from Spatio-Temporal MEG Data". *IEEE Trans. Biomed. Eng.*, Vol.BME-39(No.6):pp.541–557, June 1992.

[8] J. C. Mosher, M. E. Spencer, R. M. Leahy, and P. S. Lewis. "Error bounds for EEG and MEG dipole source localization". *Electroenceph. clinical Neurophys.*, Vol.86:pp.303–321, 1993.

[9] W. W. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 2nd ed., 1992.

[10] M. Sun, S. Qian, X. Yan, S. B. Baumann, X.-G. Xia, R. D. Dahl, N. D. Ryan, and R. J. Sclabassi. "Localizing Functional Activity in the Brain through Time-Frequency Analysis and Synthesis of the EEG". *Proc. IEEE*, Vol.84(No.9):pp.1302–1311, September 1996.

[11] S. Weiß and U. Hoppe. "Recognition and Reconstruction of Late Auditory Evoked Potentials Using Wavelet Analysis". In *Proc. IEEE Int. Symp. TFTS Analysis*, pp.473–476, Paris, 1996.