An Ensemble De-Noising Method for Spatio-Temporal EEG/ MEG Data *

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

[1] Signal Processing Division, University of Strathclyde Glasgow, UK

[2] Signal and Image Processing Institute University of Southern California, Los Angeles, CA

[3] Biophysics Group, Los Alamos National Laboratory, NM

May 19, 1998

*This work has been supported by a scholarship from William Garven Research Bequest, Glasgow, and by the National Institute of Mental Health under Grant No. MH53213. It was first presented at the 2nd IEEE UK Symposium on Applications of Time-Frequency and Time-Scale Analysis, University of Warwick, England, in August 1997.

Address for correspondence: Stephan Weiß, Signal Processing Division, Dept. of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XW, Glasgow, UK. Tel. ++44-141-548-2684, FAX ++44-141-552-2487, e-mail weiss@spd.eee.strath.ac.uk

Abstract. EEG/MEG are important tools for non-invasive medical diagnosis and basic studies of the brain and its functioning, but often applications are limited due to a very low SNR in the data. Here, we present a discrete wavelet transform (DWT) based de-noising method for spatio-temporal EEG/MEG measurements collected by a sensor array. A robust threshold selection can be achieved by incorporating spatial information and pre-stimulus data to estimate signal and noise energies. Further improvement can be gained by applying a translation-invariant approach to the derived de-noising scheme. In simulations, the performance of the proposed method is evaluated in comparison to standard de-noising and low-rank approximation, which offers some complementarity to our approach.

Keywords: EEG/ MEG, source localization, de-noising, low-rank approximation, discrete wavelet transform.

1 Introduction

The electro- and magnetoencephalogram (EEG/ MEG) are recordings of the scalp potential and magnetic field outside the head resulting from electrical neural activity in the brain. The human EEG was first recorded back in 1924, and has, together with MEG, acquired an important role as a diagnostic tool in medicine and brain research. Applications include detection of epileptic seizure [1], identification of evoked potentials in reaction to stimuli, which can be used, for example, for an objective audiogram [2], exploration of sleep states [3], or general analysis to find areas of neural activity from event related data by solving an inverse problem using the EEG/MEG measurements [4].

Our main interest has been in localizing neural sources in the brain. Transient neural current sources associated with event related EEG/ MEG are generally assumed focal in nature and can be approximated using equivalent current dipoles. Therefore using measurements from a sensor array, the location, orientation, and time series of a number of equivalent dipoles can be determined by solving an inverse problem [4]. However, even after stimulus-locked averaging from multiple trials, usually the data 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 [5], and also limits the utility of EEG/MEG in other previously mentioned applications. Here we address the problem of improving the SNR of recorded EEG/MEG data.

Since the signals of interest are transient in nature, simple bandpass filtering is not a suitable technique for separation of signal and noise. When the signals of interest are a linear combination of a small number of linearly independent time series, noise reduction can be achieved in sensor array data by exploiting the spatial structure of the underlying signals. This structure can be revealed by applying a singular value decomposition (SVD) to the data matrix and performing a low-rank approximation (LRA), thus truncating noisy unstructured contributions. LRA is implicitly performed on the data when using sub-space based source localization algorithms [4].

A different approach to improving the SNR of recorded EEG/ MEG data is to use wavelet de-noising, which is a time-frequency method to recover an unknown transient signal from broadband noise [6]. By applying a suitable orthonormal transform to the time-dimension of the data, the noise will remain spread across the transform space while the signal of interest can be parameterized by few transform coefficients that will stand out from the noise. The localizing properties of the discrete wavelet transform (DWT), in both time and frequency, make it suitable for parameterizing transient evoked responses within the EEG, as demonstrated in, for example, [7, 8]. 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 EEG/MEG problem by exploiting the additional information specific to our model of the data, and present simulation results to demonstrate the benefit of this approach to noise reduction.

The paper is organized as follows. In Section 2 we introduce a signal model for EEG/MEG based on focal neural current sources, and briefly review low-rank approximation and wavelet de-noising. The customization of the de-noising approach for our problem is the subject of Section 3. We present comparisons to other methods and results in Section 4. Our notation uses plain italicized text for scalar values, and bold face lower and upper case for vector and matrix quantities, respectively.

2 Preliminaries

This section describes the signal model for array-recorded EEG/MEG measurements, and introduces two basic concepts of noise reduction. The first one, low rank approximation, exploits the fact that the signal of interest has some spatial structure while the noise is weakly structured, if at all. The second approach searches for feature parameters in the time-frequency plane that stand out from the noise, which is subsequently

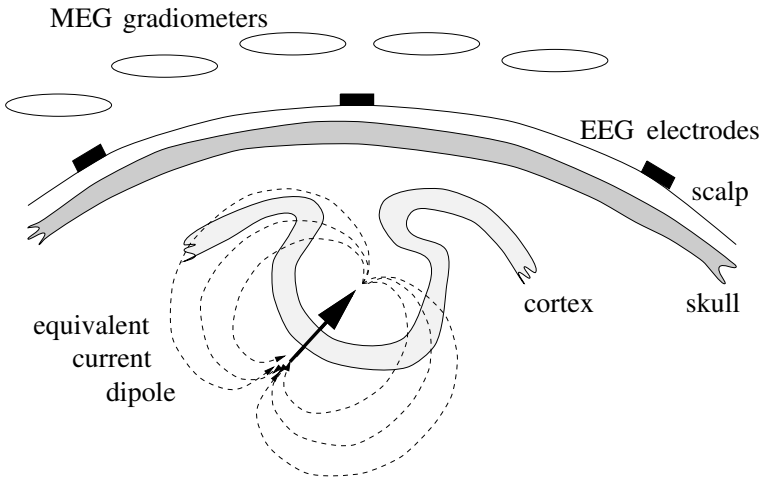

Figure 1: Modelling focal neural activity in the brain by a current dipole.

suppressed by masking or shrinking of the noise-only coefficients.

2.1 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 element sensor array is formed by the superposition of the fields from each of the r 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-L spatio-temporal data matrix

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

with
$\mathbf{G}(\mathbf{r}_i) \in \mathbb{R}^{M\times 3}$ gain matrix,
$\mathbf{q}_i \in \mathbb{R}^{3\times 1}$ dipole orientation,
$\mathbf{s}_i \in \mathbb{R}^{L\times 1}$ dipole time series, and
$\mathbf{N} \in \mathbb{R}^{M\times L}$ additive noise,

where L is the number of collected time slices. The vector $\mathbf{s}_i$,

$$\mathbf{s}_i = \begin{bmatrix} s_i[0] & s_i[1] & \cdots & s_i[L-1] \end{bmatrix}^T \tag{2}$$

contains the amplitude (or "strength") of the ith dipole over a time interval of L discrete samples, often also referred to as activation function. This strength is given a spacial direction by the fixed dipole orientation $\mathbf{q}_i$. Finally, the matrix $\mathbf{G}(\mathbf{r}_i)$ contains the gain

factors from the spatial components of the ith dipole at location $\mathbf{r}_i$ to the M sensors, which measure either electric potential in the case of EEG or the magnetic flux in the case of MEG. This so called forward model $\mathbf{G}(\mathbf{r}_i)$ can be calculated based on a model of the head accounting for the electrical properties of different layers (e.g. brain matter, scull bone, and scalp as indicated in Fig. 1). With respect to the sensor number M, current whole head MEG and EEG systems provide in the range of 64 to more than 300 channels.

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:

$$\|\mathbf{F}\|_F^2 \approx \|\mathbf{X}\|_F^2 + \|\mathbf{N}\|_F^2 \quad . \tag{3}$$

The noise power can be estimated from pre-stimulus data, corresponding to a period of $\tilde{L}$ time slices before any event-related signals are produced, yielding

$$\|\mathbf{N}\|_F^2 \approx \frac{L}{\tilde{L}} \|\tilde{\mathbf{N}}\|_F^2 \quad . \tag{4}$$

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

$$\mathrm{SNR} = \frac{\|\mathbf{X}\|_F^2}{\|\mathbf{N}\|_F^2} \tag{5}$$

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

2.2 Low-Rank Approximation

The rank of a matrix can be revealed by performing a singular value decomposition (SVD):

$$\mathbf{U}^T \mathbf{F} \mathbf{V} = \begin{bmatrix} \mathbf{\Sigma} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{F} \in \mathbb{R}^{M \times L},\ M \leq L \tag{6}$$

where $\mathbf{U} \in \mathbb{R}^{M \times M}$ and $\mathbf{V} \in \mathbb{R}^{L \times L}$ are orthogonal matrices and $\mathbf{\Sigma} = \mathrm{diag}\{\sigma_m\} \in \mathbb{R}^{M \times M}$ holds the singular values of $\mathbf{F}$ ordered as $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_M \geq 0$. The number of non-zero singular values then determines the rank of $\mathbf{F}$, which, for noiseless data, gives the number of linearly independent sources contributing to the measurements. However, in the presence of noise, the matrix will become full rank. For the case of spatially and temporally uncorrelated Gaussian noise with variance σ_{nn}^2, the last $M-r$ singular values will take the value of $\sigma_m = \sqrt{L \cdot \sigma_{nn}^2}, \forall m \in [r+1, M]$, while the first r singular values will be elevated above this noise floor by the singular values of the noiseless data matrix.

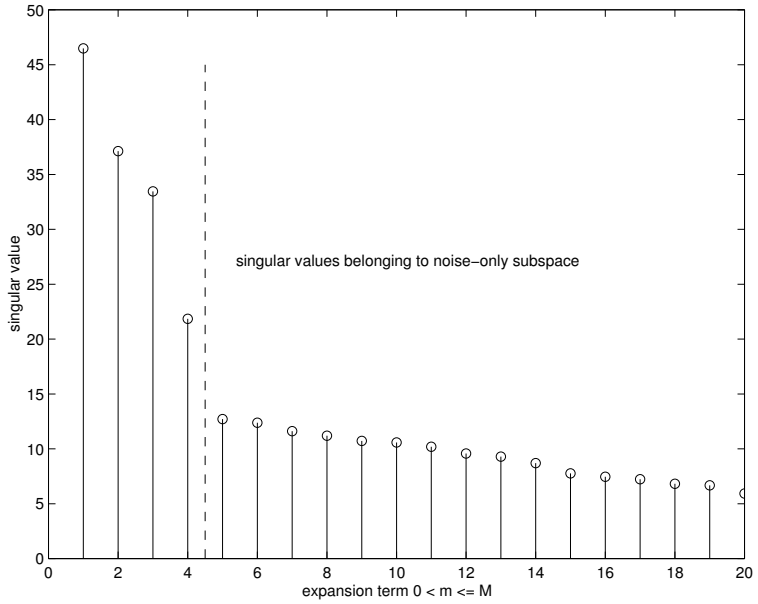

Figure 2: Example of the singular values of a data matrix for $r = 4$ independent sources measured using an array of $M = 20$ sensors and corrupted by Gaussian white noise; the slope in the singular values corresponding to the noise-only subspace is due to finite-sample statistics.

An example for a Gaussian white noise corrupted data matrix $\mathbf{F} \in \mathbb{R}^{20\times 100}$ containing linearly combined observations of $r = 4$ temporally independent sources is shown in Fig. 2.

Using an SVD, $\mathbf{F}$ can be written in the expansion form [9]

$$\mathbf{F} = \sum_{m=1}^{M} \sigma_m \mathbf{u}_m \mathbf{v}_m^T \tag{7}$$

where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots \mathbf{u}_M] \in \mathbb{R}^{M\times M}$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_L] \in \mathbb{R}^{L\times L}$ span the column and row space of $\mathbf{F}$. If $\mathbf{F}$ contains structured data emanating from r linearly independent sources in the presence of independent and identically distributed (iid) white noise, the expansion (7) can be truncated after r terms thus suppressing the noise introduced by the remaining $M - r$ non-zero singular values. We refer to the truncated form

$$\tilde{\mathbf{F}}_{\mathrm{LRA}} = \sum_{m=1}^{r} \sigma_m \mathbf{u}_m \mathbf{v}_m^T \tag{8}$$

as a low-rank approximation of $\mathbf{F}$. It can also be seen as a projection of $\mathbf{F}$ onto its signal-subspace spanned by $\mathbf{U}_r = [\mathbf{u}_1, \cdots \mathbf{u}_r]$, $\tilde{\mathbf{F}}_{\mathrm{LRA}} = \mathbf{U}_r \mathbf{U}_r^T \mathbf{F}$, which eliminates any components of $\mathbf{F}$ in the noise-only subspace $\mathbf{U}_r^{\perp}$, where $(\cdot)^{\perp}$ refers to the orthogonal complement. Thus, for white Gaussian noise, the noise variance is reduced by a factor r/M.

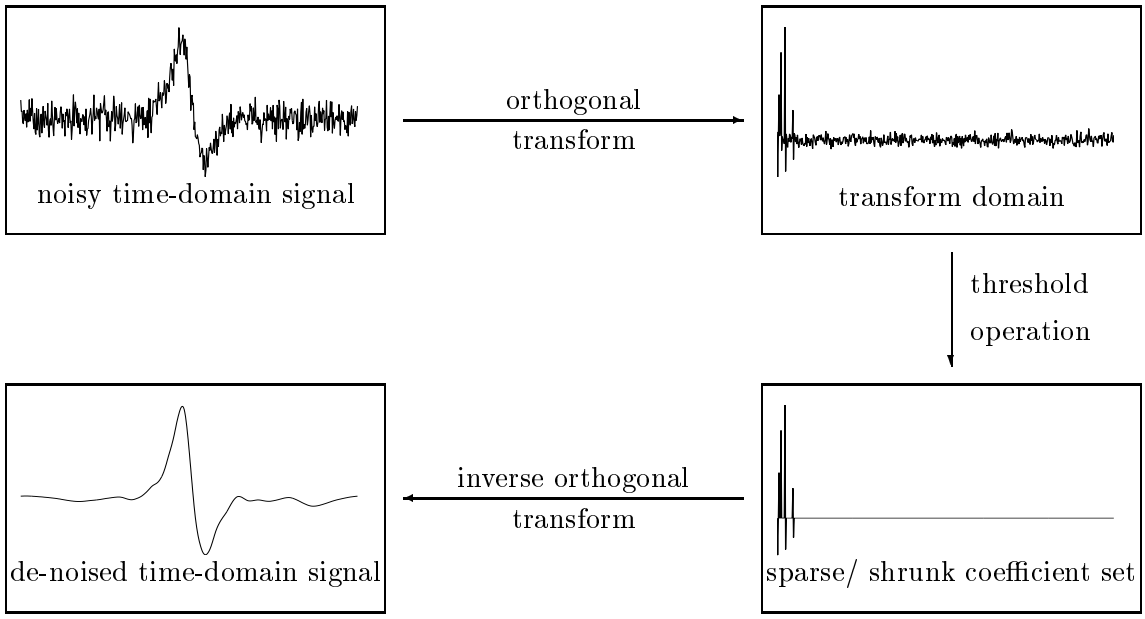

Figure 3: De-noising of a one dimensional time-series: in the transform domain, additive noise remains smeared, while the signal of interest is parameterized by only few coefficients. By thresholding and inverse transformation, a noise reduction can be achieved.

2.3 De-Noising

De-noising as originally proposed by Donoho and Johnstone [6] is applicable to 1-dimensional signals corrupted by white noise. The signal is transformed, the transform coefficients thresholded according to some criterion, and an inverse transform used to obtain a noise reduction. This procedure is shown in Fig. 3. De-noising relies on an appropriate transform that is able to parameterize the signal of interest with a few significant coefficients while, for an orthogonal transform, the noise will remain spread over the entire transform domain.

A discrete wavelet transform (DWT) is usually employed as its ability to yield local representation in both time and frequency domain is advantageous for the analysis of transient signals. The quality of the parameterization determines the ability to recover a minimally distorted de-noised signal. If we assume iid Gaussian noise, a parameterization by $K \leq L$ coefficients yields a noise reduction of approximately K/L. A huge variety of wavelets has been investigated in literature possessing different properties such as minimum length of support and maximum smoothness [10], near symmetry, or good localization in both time and frequency [11].

A second important choice is the thresholding method. Hard-thresholding zeroes every coefficient that falls below a defined threshold, while for soft-thresholding, a continuous non-linear function is applied to the transform coefficients. The most straightforward approach is to additionally shrink all surviving coefficients by the amount of the threshold. Threshold selection is determined by whether the de-noising procedure is optimized in a mean squared error, minimax, or visually appealing ("visu-shrink" [12]) sense [6]; often the choice is heuristic. Additionally, different threshold optimizations can be combined with adaptive and transform-level dependent schemes [13].

The threshold schemes generally assume white Gaussian iid noise with unit variance corrupting the signal of interest. Therefore, prior to de-noising, the signal has to be scaled to set the noise variance correctly. Often, the initial noise variance is estimated using the highpass part of the data, i.e. the wavelet coefficients at the finest transform level, and a normalization with their median absolute deviation is applied. This method assumes a sufficient smoothness of the underlying signal of interest such that only noise is present at high frequencies [12].

3 Ensemble De-Noising

This section will introduce a particular de-noising scheme, in which we exploit prestimulus data to estimate noise and signal energies. We add security and a degree of determinism to the threshold selection by incorporating the spatial dimension of the data in the decision, which provides ensembles probes of the noise process. Hence we refer to our method as ensemble de-noising (EDN).

3.1 Concept

We represent a real valued DWT by an orthogonal transform matrix $\mathbf{T}$, such that $\mathbf{y} = \mathbf{T}\mathbf{x}$ is the DWT of $\mathbf{x}$ [14]. Applying the transform to the temporal dimension of the data matrix $\mathbf{F}$, we can express the de-noising procedure as

$$\tilde{\mathbf{F}} = \mathbf{\Theta}\big(\mathbf{F}\mathbf{T}^T\big)\,\mathbf{T}, \tag{9}$$

where $\mathbf{\Theta}(\cdot)$ performs a threshold operation 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: since we expect to see similar characteristics in the transform coefficients across the sensor array [3], we can use a mask common to the whole array rather than de-noising each measurement time series separately, i.e.

$$\tilde{\mathbf{F}}_{\mathrm{EDN}} = \mathbf{F} \underbrace{\mathbf{T}^T\mathbf{M}\mathbf{T}}_{\text{TF-filter}} \tag{10}$$

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

$$\mu_l = \begin{cases} 1 & : \quad \bar{t_f^2}(l) \geq \theta \\ 0 & : \quad \bar{t_f^2}(l) < \theta \end{cases} \qquad\qquad l = 0(1)L-1 \tag{11}$$

depending on the squared transform coefficients averaged over the spatial dimension

$$\bar{t_f^2}(l) = \frac{1}{\|\mathbf{F}\mathbf{T}^T\|_F^2} \sum_{m=0}^{M-1} |t_f(m,l)|^2 . \tag{12}$$

The reason for averaging over the energy of the coefficients rather than the coefficient values themselves lies in the forward model, which linearly combines the source transients and therefore sign changes across the array must be considered. The normalization $\|\mathbf{F}\mathbf{T}^T\|_F^2$ in (12) is to ensure $\|\bar{\mathbf{t}_f^2}\|_1 = 1$, i.e. the overall energy represented in the parameters $\bar{t_f^2}(l)$ is unity. With (3), this energy can be separated into contributions form signal of interest and noise.

3.2 Threshold Selection

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 (5). The elements of $\bar{\mathbf{t}_f^2}$ are re-ordered by a permutation matrix $\mathbf{P}$ such that the resulting vector

$$\tilde{\mathbf{t}_f^2} = \mathbf{P}\bar{\mathbf{t}_f^2} \tag{13}$$

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

$$s(l) = \sum_{i=0}^{l} \tilde{t_f^2}(i), \tag{14}$$

represents the normalized energy in the l largest coefficients. We determine the number

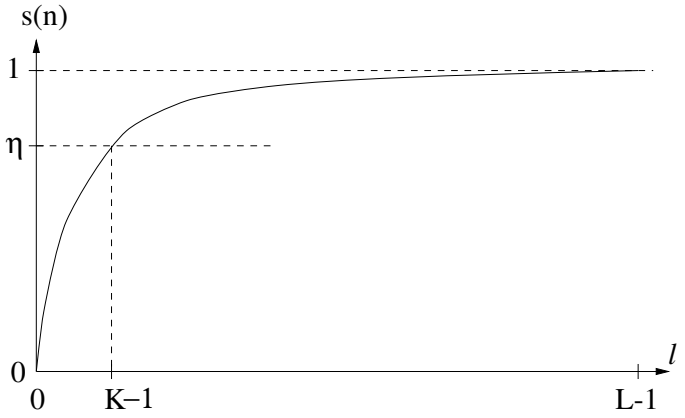

Figure 4: Threshold selection for ensemble de-noising.

of coefficients, K, to retain by comparing $s(l)$ to the ratio η of energies of $\mathbf{X}$ and $\mathbf{F}$ as illustrated in Fig. 4. Therefore, η gives the fraction for the contribution of the signal of interest to the overall energy contained in the measured data. 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 orthonormal transform matrix $\mathbf{T}$. The ratio η can be calculated from the EEG/ MEG data using assumptions (3) and (4)

$$\eta = \frac{\|\mathbf{X}\|_F^2}{\|\mathbf{F}\|_F^2} \approx \frac{\|\mathbf{F}\|_F^2 - \frac{L}{\tilde{L}}\|\tilde{\mathbf{N}}\|_F^2}{\|\mathbf{F}\|_F^2}. \tag{15}$$

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

By evaluating (11), and insertion into (10) the noise in the data matrix $\mathbf{F}$ has been reduced. We refer to this method as ensemble de-noising (EDN), since the spatial dimension of the data presents us with ensemble probes of the noise process.

How does the ensemble strategy in (12) influence the probability of picking the correct coefficients in noise? For the analysis, we assume as a simplification that the signal of interest can be represented by a single coefficient, t_x, having constant modulus across the array, and that the noise coefficients $t_{l,i}$ are samples of a Gaussian iid process $\mathcal{N}$ with zero mean and variance σ_{nn}^2. We are interested in the probability

$$P(z \geq M \cdot t_x^2), \quad \text{with } z = \sum_{i=0}^{M-1} t_{l,i}^2, \quad t_{l,i} \in \mathcal{N}(0, \sigma_{nn}^2) \tag{16}$$

as a measure for the probability that the signal will be hidden in noise. The specific threshold in (16) refers to the case where signal and noise have identical power. Thus, the random variable z is characterized by an appropriately scaled chi-square distribution with M degrees of freedom [15].

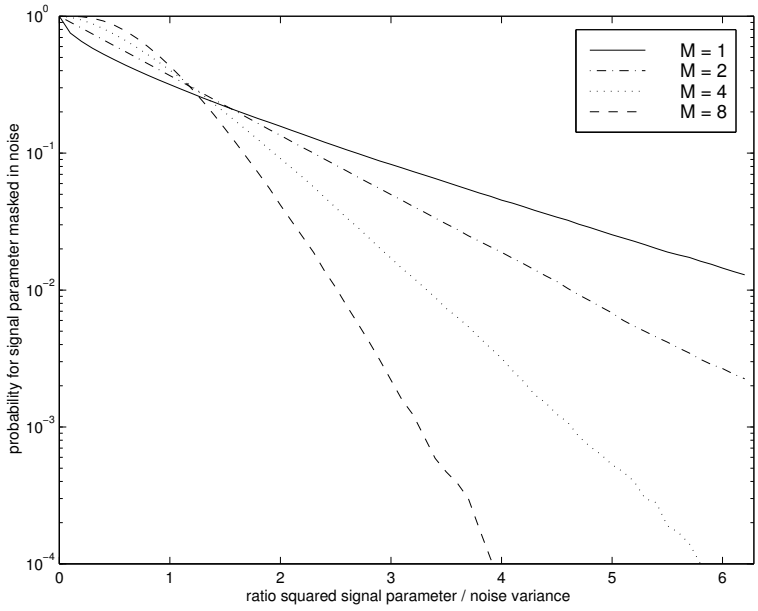

Figure 5: Statistical security of looking across the sensor array in dependency of the ratio between feature parameter and variance of corrupting noise, and the dimensionality M of the sensor array.

Fig. 5 shows the probability $P(z \geq M \cdot t_x^2)$ drawn over the ratio t_x^2/σ_{nn}^2 for different sensor array dimensions M. It can be seen that for a reasonable ratio t_x^2/σ_{nn}^2, averaging the coefficient energy across the M-array considerably enhances the statistical security for picking the correct coefficient. Note, that for a spatio-temporal data matrix of L recorded discrete time slices and parameterization of the signal of interest by a single coefficient t_x,

- the SNR of the data is given by $t_x^2/(L \cdot \sigma_{nn}^2)$; and
- the probability that t_x is selected despite the presence of L noisy coefficients yields $(1 - P(z \geq M \cdot t_x^2))^L$.

The high error probability $P(z \geq M \cdot t_x^2)$ for small ratios t_x^2/σ_{nn}^2 in Fig. 5 highlights both the fact that de-noising is unsuitable for very low SNR and the necessity that the transform in the de-noising process sufficiently parameterizes the signal of interest.

3.3 Translation-Invariant Ensemble De-Noising

As mentioned earlier, it is implicitly assumed in the de-noising procedure that the signals of interest, i.e. the dipole time series, can be represented by a small subset of the basis

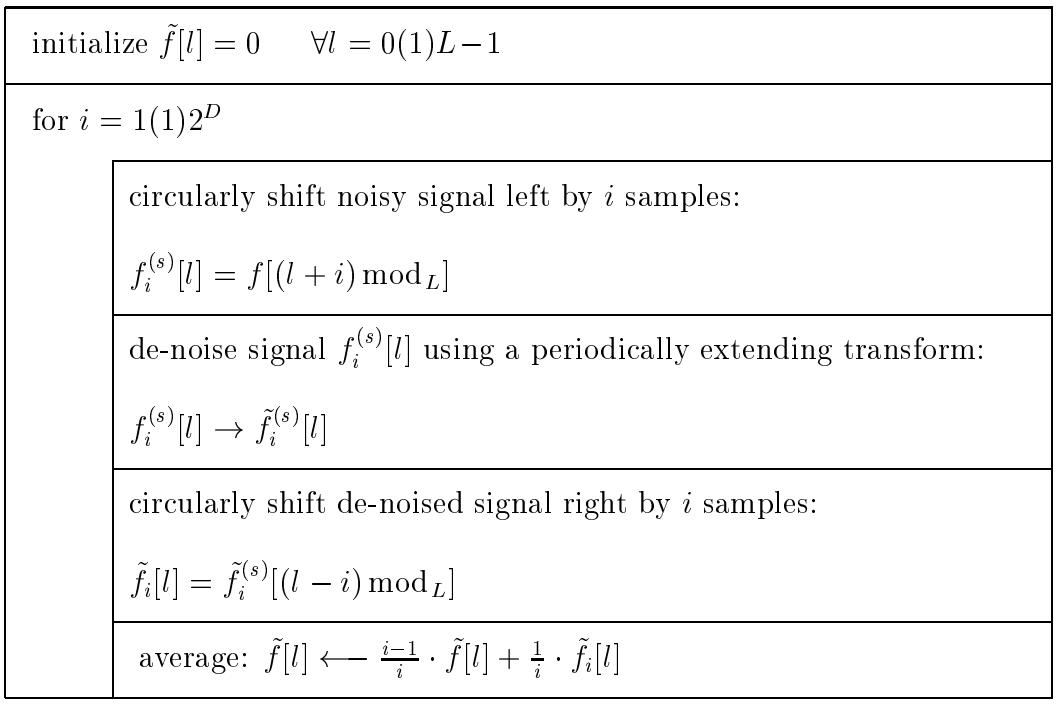

Figure 6: Flow-chart for translation-invariant (TI) de-noising of a 1-d time series $\mathbf{f}$. Variables with superscript $^{(s)}$ refer to circularly shifted signals, D is the number of levels of the applied DWT. For TI-EDN circular shift operations are applied to all rows of the data matrix $\mathbf{F}$.

functions of the transform. However, problems also arise due to the phase-sensitivity through the cyclo-stationarity of the DWT associated with the decimation process. As a result, the basis functions sit orthogonal to each other on a fixed dyadic grid. Therefore even if the signal of interest matches one of the basis functions, a phase shift of the signal destroys the representation by a single coefficient and can blur the signal energy over several transform coefficients of smaller amplitude, which may be masked in the noise and get subsequently truncated when falling below the threshold selected according to Sec. 3.2.

These truncations result in oscillations similar to Gibbs phenomena in the reconstructed de-noised time-series [16]. A translation-invariant (TI) approach recently discussed in literature [16, 17, 18, 19] can address this problem. We have followed the approach of [16], where shifted versions of the data matrix are ensemble de-noised, back-shifted, and averaged as shown in the flow chart in Fig. 6. This presents a compromise over searching for an optimum shift with minimum entropy in the transform domain, i.e. energy concentration in as few coefficients as possible [20], as a shift that is optimal for one transient may not be optimal for another transient contained in the same data.

The shift operation in Fig. 6 is executed circularly. As the sensor time series forming the rows of the data matrix $\mathbf{F}$ are finite intervals, the DWT — efficiently evaluated using an octave filter bank [11] rather than the matrix multiplication in (10) — requires an extension of the signal. If a periodic or symmetric extension is chosen, filtering can be performed in steady-state, thus avoiding transients that distort the subband signals holding the DWT coefficients [8, 21]. Although a periodic extension generally suffers from discontinuities and "wrap-around" effects at the margins of the temporal interval [14], it has been employed for this work as the TI approach can be easily realized by circular shifts. In terms of implementation, it is not required to calculate a full transform for every shift from scratch, as reduced computational schemes can be applied [12, 22, 17, 23].

4 Simulations and Results

We compare the de-noising methods described in Sections 2.3 and 3 to the low rank approximation of the data matrix obtained via a truncation of the SVD expansion as described in Section 2.2. To evaluate the results in terms of noise reduction, we have to use synthetic data. As source transients $\mathbf{s}_i$, superpositions of Hermite-type functions have been employed, which are derivatives of Gaussian window functions and realistic for simulating activations [24]. Examples for such source transients are given in Fig. 7. The wavelet transform uses a Symmlet-8 wavelet [12, 10], which has a relatively good frequency selectivity due to maximum smoothness and gives an approximate alignment of non-stationarites in various scales due to its near-symmetry. However, the choice of the 8-coefficient wavelet is heuristic and based on empirical results with wavelet filters of different lengths in combination with the realistic simulation data employed in the following.

Fig. 8 shows the averaged results over 25 trials with different noise power spectral densities (PSD) and realistic source data for different de-noising techniques and the low rank approximation. The figure shows the final SNR $= 10 \cdot \log_{10}(\|\mathbf{X}\|_F^2 \cdot \|\tilde{\mathbf{F}} - \mathbf{X}\|_F^{-2})$ versus the cut-off frequency of the coloured (lowpass) noise. The initial SNR of the data matrix $\mathbf{F}$ was 5dB. The temporal colouring of the noise is achieved by lowpass filtering, with the cut-off frequency indicated on the abscissa. For standard de-noising of each sensor time series, we achieved best results using a visu-shrink soft threshold [12]. The SNR produced by ensemble de-noising is consistently higher due to exploitation of pre-stimulus and spatial information. However, note that further improvements are achieved

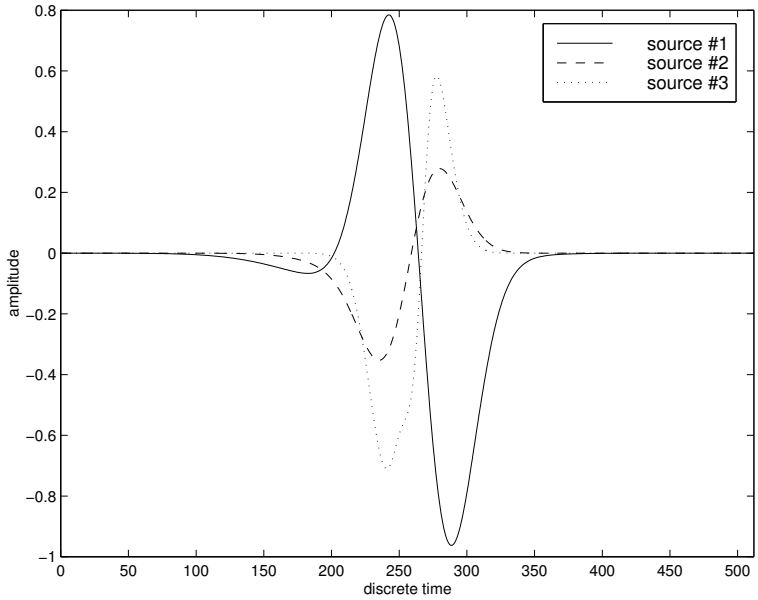

Figure 7: Example time series $s_i[l]$, $i = \{0, 1, 2\}$, to simulate three source activations consisting of superposed Hermite-type functions, for construction of synthetic EEG/MEG data matrices.

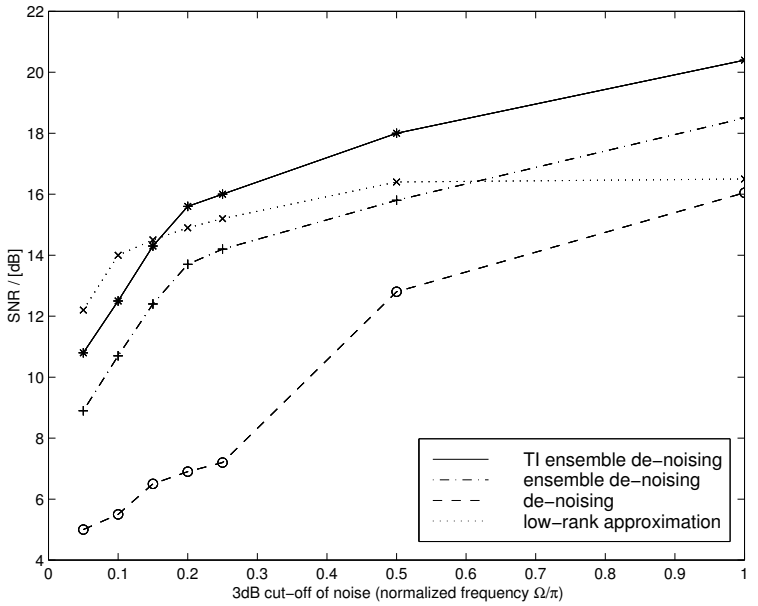

Figure 8: Final SNR achieved by different de-noising methods applied to data from a 73-sensor array. Initial SNR is 5dB, with temporally coloured noise of different bandwidths produced using a 5th order Butterworth filter. The signal of interest arises from three dipolar sources activated by Hermite-type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

by applying translation-invariant (TI) ensemble de-noising.

Example time series for de-noising results in Fig. 8 are given in Figs. 9–11. Fig. 9 shows the measurement at a specific sensor, i.e. one row of the data matrix, as a linear combination of the source transients in Fig. 7, which is corrupted by noise at 5dB SNR, having a PSD with a normalized angular cut-off frequency at $\Omega/\pi = 0.15$. The standard de-noising solution using a visu-shrink threshold is drawn as a solid line, and clearly suffers from the strong colouring of the added noise. The reconstructed measurements using various noise-reduction techniques are shown in Fig. 10, with the error between the noiseless time series and the reconstructed response given in Fig. 11. A remarkable difference between the proposed ensemble de-noising methods and low-rank approximation is that the time-frequency approach well suppresses noise outside the activation interval of the sources, within which low rank approximation appears to give a better fit to the original signal.

To further investigate the suspected complementarity between ensemble de-noising and the SVD based noise reduction, we look at an extreme simulation, where a single source transient is matched by the analysis function (again a Symmlet-8) of the DWT. Therefore in the resulting situation, ensemble de-noising can be expected to give its maximum benefit. The results in terms of noise-reduction in the reconstructed matrix $\tilde{\mathbf{F}}$, are displayed in Fig. 12 for different initial SNRs in the data matrix $\mathbf{F}$. 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, as indicated in Section 3.2. For high SNR, both methods reach a constant level, since the noise reductions approach the factors $(M-r)/M$ for low-rank approximation (LRA) and $(L–K)/L$ for de-noising, as discussed in Sections 2.2 and 2.3. The partial complementarity of LRA and de-noising becomes evident from a combination of both methods, where the data is first ensemble de-noised and then low-rank approximated. The result of this combined operation is drawn as dotted line in Fig. 12, with increased noise reduction over either method separately. Since subspace-based dipole source localization algorithms inherently perform a low-rank approximation [4], ensemble de-noising is expected to give additional benefit.

For the situation in Fig. 12, Tab. 1 contains results of source parameter estimation [4] for different SNRs with and without ensemble de-noising (EDN). Performance measures are the average of the mislocation of the dipolar source, $\|\tilde{\mathbf{r}} - \mathbf{r}\|_2$ (in cm), and the relative error of the reconstructed source activation function, $\|\tilde{\mathbf{s}} - \mathbf{s}\|_2$ with $\|\mathbf{s}\|_2 = 1$, where tilded variables refer to estimated quantities. There is some minor improvement in estimating the source location and slight enhancement of the reconstructed time series

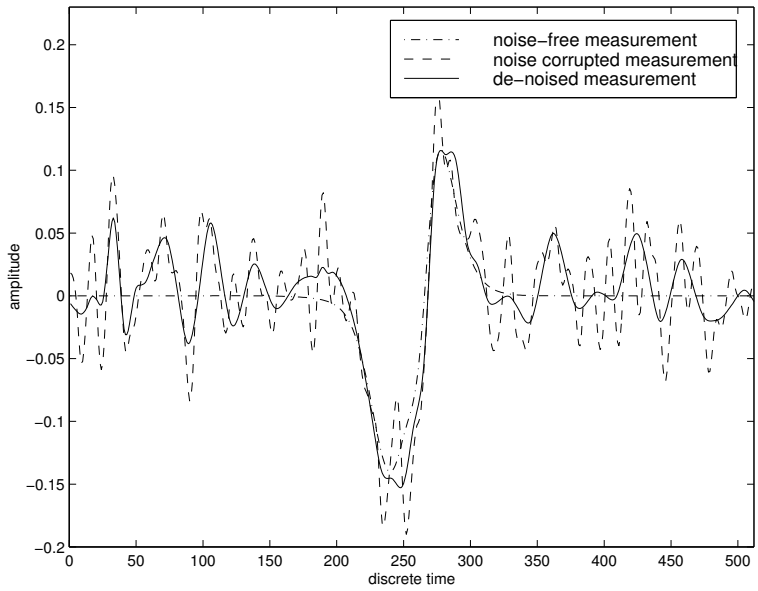

Figure 9: Linear combination of source transients in Fig. 7 measured at one sensor (dash-dotted); with coloured Gaussian noise added at 5dB SNR (dashed); and reconstructed time-series using standard de-noising.

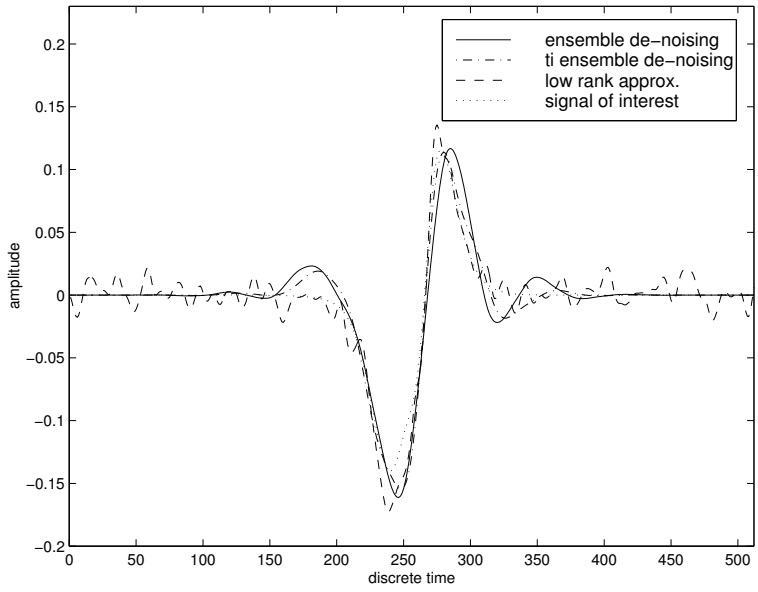

Figure 10: Reconstructed time series of sensor measurement in Fig. 9 applying various noise reduction methods.

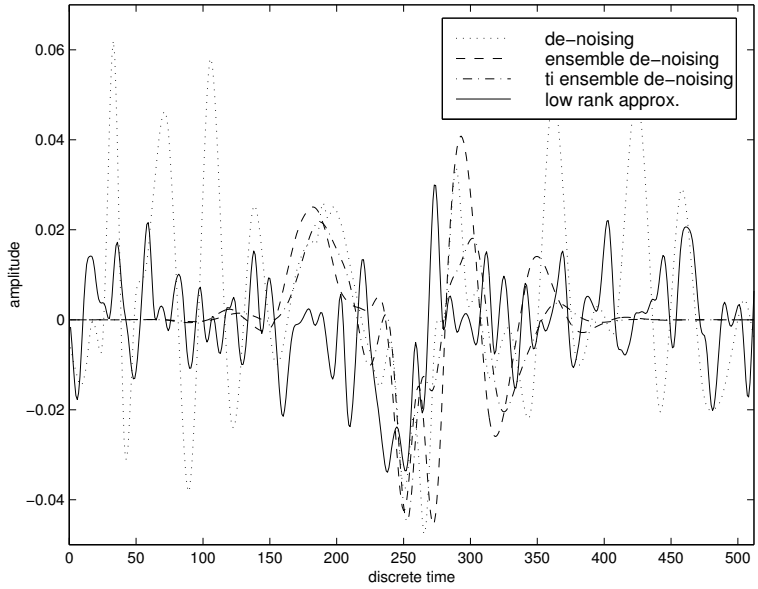

Figure 11: Errors between the reconstructed time series in Fig. 10 and the underlying signal of interest in Fig. 9.

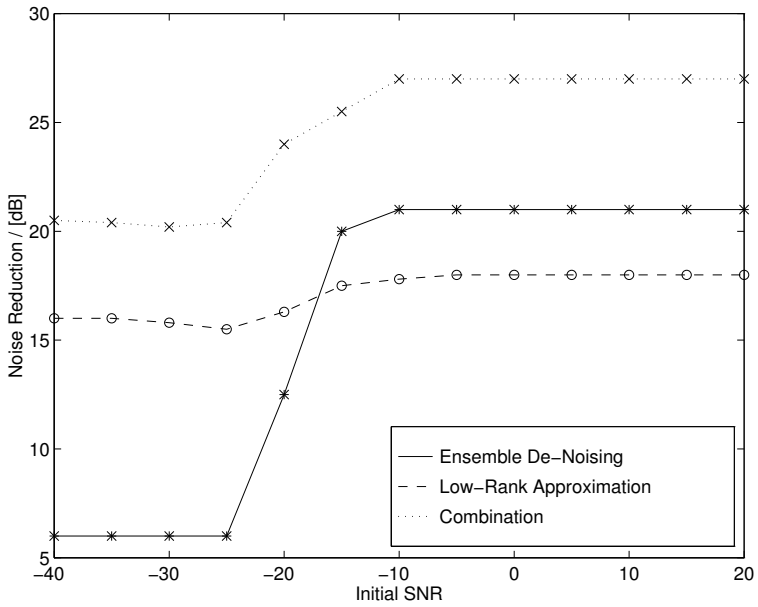

Figure 12: Comparison of combination of ensemble de-noising with a low-rank approximation for different SNRs.

Table 1: Averaged deviation of location (in cm) and relative time series error for source parameter estimation according to [4] for a single dipole.

SNR	EDN	$\overline{\|\Delta\mathbf{r}\|_2}$	$\overline{\|\Delta\mathbf{s}\|_2}$
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

when applying the denoising method prior to source localization.

5 Conclusion

We have introduced a wavelet de-noising method for spatio-temporal EEG/MEG data, which incorporates pre-stimulus and spatial information in the selection of the threshold. The resulting scheme yields an enhanced SNR improvement over standard de-noising, and offers some complementarity to low-rank approximation, a noise reduction technique based on an SVD of the data rather than a time-frequency analysis as performed with de-noising. 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 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.* [20] 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 is also the possibility of using soft thresholding [25] in (10), with an appropriate threshold function designed to preserve the estimated energy of the signal of interest.

References

[1] Wendling F, Bellanger JJ, Badier JM, Coatrieux JL. Extraction of Spatiotemporal Signatures from Depth EEG Seizure Signals Based on Objective Matching in Warped Vectorial Observations. IEEE Trans Biomed Eng 1996; 43(10):990–1000, 1996.

[2] Hoppe U, Eysholdt U, Weiß S. A Sequential Detection Method for Late Auditory Evoked Potentials. In Proc Intern Conf IEEE Engineering in Medicine and Biology Society, Amsterdam, The Netherlands, November 1996.

[3] Sun M, Qian S, Yan X, et al. Localizing Functional Activity in the Brain through Time-Frequency Analysis and Synthesis of the EEG. Proc IEEE 1996; 84(9):1302–1311.

[4] Mosher JC, Lewis PS, Leahy RM. Multiple Dipole Modeling and Localization from Spatio-Temporal MEG Data. IEEE Trans Biomed Eng 1992; BME-39(6):541–557.

[5] Mosher JC, Spencer ME, Leahy RM, Lewis PS. Error bounds for EEG and MEG dipole source localization. Electroenceph clinical Neurophys 1993; 86:303–321.

[6] Donoho DL, Johnstone IM. Ideal Spatial Adaptation Via Wavelet Shrinkage. Biometrika 1994; 81:425–455.

[7] Bartnik E, Blinowska K, Durka P. Single Evoked Potential Reconstruction by Means of Wavelet Transform. Biol Cybern 1992; 67:175–181.

[8] Weiß S, Hoppe U. Recognition and Reconstruction of Late Auditory Evoked Potentials Using Wavelet Analysis. In Proc IEEE Intern Symp Time-Frequency Time-Scale Anal 1996; pp.473–476, Paris, France.

[9] Golub GH, Van Loan CF. Matrix Computations. 3rd edn. Johns Hopkins University Press, Baltimore, Maryland 1996.

[10] Daubechies I. Ten Lectures on Wavelets. SIAM, Philadelphia 1992.

[11] Mallat SG. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. IEEE Trans Pattern Anal Machine Intelligence 1989; 11(7):674–692.

[12] Buckheit J, Chen S, Donoho DL, Johnstone IM, Scargle J. Wavelab 7.01. Available from http://playfair.stanford.edu/~wavelab, 1996. Stanford University.

[13] Donoho DL, Johnstone IM. Adapting to Unknown Smoothness Via Wavelet Shrinkage. J American Statistical Association 1995; 90(432):1200–1224.

[14] Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical Recipes in C. 2nd edn. Cambridge University Press, Cambridge 1992.

[15] Papoulis A. Probability, Random Variables, and Stochastic Processes. 3rd edn. McGraw-Hill, New York 1991.

[16] Coifman RR, Donoho DL. Translation-Invariant De-Noising. In: A. Antoniadis (ed). Wavelets and Statistics. Springer Verlag 1995.

[17] Liang J, Parks TW. A Translation Invariant Wavelet Representation Algorithm with Applications. IEEE Trans Signal Processing 1996; 44(2):225–232.

[18] Pesquet JC, Krim H, Carfantan H. Time-Invariant Orthonormal Wavelet Representations. IEEE Trans Signal Processing 1996; 44(8):1964–1970.

[19] Sari-Sarraf H, Brzakovic D. A Shift-Invariant Discrete Wavelet Transform. IEEE Trans Signal Processing 1997; 45(10):2621–2626.

[20] Coifman RR, Wickerhauser MV. Entropy-Based Algorithms for Best Basis Selection. IEEE Trans Inf Theory 1992; 38(2):713–718.

[21] Strang G, Nguyen T. Wavelets and Filter Banks. Wellesley–Cambridge Press, Wellesley, MA 1996.

[22] Beylkin G. On the Representation of Operators in Bases of Compactly Supported Wavelets. SIAM J Numerical Analysis 1992; 29:1716–1740.

[23] Delfs C, Jondral F. New Aspects on Classification Using Translation-Invariant Wavelet Packet Transforms. In Proc. 2nd IEEE UK Symp Appl Time-Frequency Time-Scale Methods. University of Warwick, England 1997. pp 73–76.

[24] Raz J, Biggins CA, Turetsky B, Fein G. Frequency Domain Dipole Localization: Extension of the Method and Applications to Auditory and Visual Evoked Potentials. IEEE Trans Biomed Eng 1993; 40(9):909–920.

[25] Donoho DL. De-Noising by Soft-Thresholding. IEEE Trans Inf Theory 1995; 41(3):613–627.

List of Figures

1 Modelling focal neural activity in the brain by a current dipole. 4

2 Example of the singular values of a data matrix for $r = 4$ independent sources measured using an array of $M = 20$ sensors and corrupted by Gaussian white noise; the slope in the singular values corresponding to the noise-only subspace is due to finite-sample statistics. 6

3 De-noising of a one dimensional time-series: in the transform domain, additive noise remains smeared, while the signal of interest is parameterized by only few coefficients. By thresholding and inverse transformation, a noise reduction can be achieved. 7

4 Threshold selection for ensemble de-noising. 10

5 Statistical security of looking across the sensor array in dependency of the ratio between feature parameter and variance of corrupting noise, and the dimensionality M of the sensor array. 11

6 Flow-chart for translation-invariant (TI) de-noising of a 1-d time series $\mathbf{f}$. Variables with superscript $^{(s)}$ refer to circularly shifted signals, D is the number of levels of the applied DWT. For TI-EDN circular shift operations are applied to all rows of the data matrix $\mathbf{F}$. 12

7 Example time series $s_i[l]$, $i = \{0, 1, 2\}$, to simulate three source activations consisting of superposed Hermite-type functions, for construction of synthetic EEG/MEG data matrices. 14

8 Final SNR achieved by different de-noising methods applied to data from a 73-sensor array. Initial 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. 14

9 Linear combination of source transients in Fig. 7 measured at one sensor (dash-dotted); with coloured Gaussian noise added at 5dB SNR (dashed); and reconstructed time-series using standard de-noising. 16

10 Reconstructed time series of sensor measurement in Fig. 9 applying various noise reduction methods. 16

11 Errors between the reconstructed time series in Fig. 10 and the underlying signal of interest in Fig. 9. 17

12 Comparison of combination of ensemble de-noising with a low-rank approximation for different SNRs. 17

Mathematical Symbols

$(\cdot)^T$	(uppercase) matrix or vector transpose
$(\cdot)^\perp$	(perp) orthogonal space
$\mathbf{0}$	(null, bold) null matrix
D	(uppercase) depth of DWT (number of iterations of halfband filter operations)
η	(greek eta, lowercase) energy ratio
EDN	(uppercase, script) subscript indicating ensemble-denoising
$\mathbf{F}$	(uppercase, bold) spatio-temporal data matrix
$\tilde{\mathbf{F}}_{\mathrm{EDN}}$	(uppercase, bold, tilde) spatio-temporal data matrix after ensemble de-noising (EDN)
$\tilde{\mathbf{F}}_{\mathrm{LRA}}$	(uppercase, bold, tilde) spatio-temporal data matrix after low rank approximation (LRA)
$\mathbf{G}$	(uppercase, bold) gain matrix
K	(uppercase) number of non-zero time samples after hard-threshold de-noising
l	(lowercase, bold) discrete time index
L	(uppercase, bold) number of time slices / samples
$\tilde{L}$	(uppercase, bold, tilde) number of time slices / samples recorded for pre-stimulus data
LRA	(uppercase, script) subscript indicating low rank approximation
M	(uppercase) number of sensors
$\mathbf{M}$	(uppercase, bold) diagonal threshold / masking matrix
μ_l	(greek mu, lowercase) masking value for lth coefficient of sensor time series
$\mathbf{N}$	(uppercase, bold) noise-only component of the spatio-temporal data-matrix
$\tilde{\mathbf{N}}$	(uppercase, bold, tilde) spatio-temporal matrix with noise-only pre-stimulus data
$\mathcal{N}(\mu, \sigma^2)$	(uppercase, calligraphy) normally distributed process with mean μ and variance σ^2
Ω	(greek Omega, uppercase) normalized (angular) frequency $\Omega = \omega T_s$, with sampling period T_s
$\mathbf{P}$	(uppercase, bold) permutation matrix
$P(z)$	(uppercase) probability of event z
$\mathbf{q}_i$	(lowercase, bold) spatial orientation of ith dipolar source
$\mathbf{r}_i$	(lowercase, bold) spatial location of ith dipolar source

r	(lowercase) number of temporally independent dipolar sources
$\mathbb{R}^{M\times L}$	(uppercase, blackboard (amstex)) set of M-by-L matrices with real entries
$s(n)$	(lowercase) cumulative sum on re-arranged averaged squared transform coefficients
$\mathbf{s}_i$	(lowercase, bold) vector holding N discrete time samples of activation function of ith dipolar source
σ_m	(greek sigma, lowercase) singular value
σ_{nn}^2	(greek sigma, lowercase) variance of noise
$\mathbf{\Sigma}$	(greek Sigma, uppercase, bold) diagonal matrix holding singular values
SNR	(uppercase, script) signal-to-noise ratio
$\mathbf{T}$	(uppercase, bold) (discrete wavelet) transform matrix
$t_f(m,l)$	(lowercase) nth transform coefficient of the mth sensor time series
$\bar{t}_f^2(l)$	(lowercase) nth squared transform coefficient averaged across the sensor array
$\bar{\mathbf{t}}_f^2$	(lowercase, bold) $\bar{t}_f^2(n)$ arranged in a vector
$\tilde{\mathbf{t}}_f^2$	(lowercase, bold) re-arranged averaged squared transform coefficient vector
$t_{l,i}$	(lowercase) transform coefficients representing noise-only components
t_x	(lowercase) noise-free transform coefficient parameterizing signal of interest
θ	(greek theta, lowercase) energy threshold
$\mathbf{\Theta}(\cdot)$	(greek Theta, uppercase, bold) matrix valued threshold function
$\mathbf{U}$, $\mathbf{V}$	(uppercase, bold) orthogonal matrices produced by singular value decomposition
$\mathbf{u}_m$, $\mathbf{v}_m$	(lowercase, bold) vectors spanning the matrices $\mathbf{U}$, $\mathbf{V}$
$\mathbf{X}$	(uppercase, bold) noise-free spatio-temporal data matrix only containing measurements due to source transients