

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Time-shift denoising source separation

Alain de Cheveigné^{a,b,*}^a Laboratoire de Psychologie de la Perception, UMR 8581, CNRS and Université Paris Descartes, France^b Département d'Etudes Cognitives, Ecole Normale Supérieure, France

ARTICLE INFO

Article history:

Received 13 January 2010

Received in revised form 24 February 2010

Accepted 1 March 2010

Keywords:

MEG

Magnetoencephalography

EEG

Electroencephalography

Noise reduction

Artifact removal

Principal component analysis

ABSTRACT

I present a new method for removing unwanted components from neurophysiological recordings such as magnetoencephalography (MEG), electroencephalography (EEG), or multichannel electrophysiological or optical recordings. A spatiotemporal filter is designed to partition recorded activity into noise and signal components, and the latter are projected back to sensor space to obtain clean data. To obtain the required filter, the original data waveforms are delayed by a series of time delays, and linear combinations are formed based on a criterion such as reproducibility over stimulus repetitions. The time shifts allow the algorithm to automatically synthesize multichannel finite impulse response filters, improving denoising capabilities over static spatial filtering methods. The method is illustrated with synthetic data and real data from several biomagnetometers, for which the raw signal-to-noise ratio of stimulus-evoked components was unfavorable. With this technique, components with power ratios relative to noise as small as 1 part per million can be retrieved.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Magnetoencephalography (MEG) and electroencephalography (EEG) measure brain activity non-invasively, but the measured signals often have an unfavorable signal-to-noise ratio (SNR) relative to environmental and sensor noise or competing physiological and brain components. Invasive techniques such as SEEG or ECoG in humans or electrode arrays in animal models offer better SNR. However, even in that case there may exist weaker components, dwarfed by the stronger components, and as the field matures they may become worthy of attention.

Multichannel sensor and electrode arrays, linear electronics, and high-resolution digital representations offer considerable leverage for multichannel signal processing techniques to resolve multiple sources and extract weak components (Parra et al., 2005; Cichocki, 2004). Independent component analysis (ICA) finds linear combinations of sensors likely to correspond to independent sources that can then be sorted between target and noise. Beamforming likewise finds a spatial filter that zeros out unwanted components based on their spatial signature, while retaining components from sources within a region of interest (Sekihara et al., 2001, 2006). A recent technique called *denoising source separation* (DSS) performs a similar operation based on criteria such as reproducibility or spectral content (Särelä and Valpola, 2005; de Cheveigné and Simon,

2008b). Beamforming, DSS and ICA are data-dependent techniques, in contrast to a method such as signal space separation (SSS) that sorts between noise and target based on geometric constraints (Taulu et al., 2005). All these techniques benefit from a large number of sensors. As sensor count is increased the improvement for *source localization* may taper off (Ahonen et al., 1993), but it continues to progress for source separation (Vrba et al., 2004).

For a given system, the number of available sensors limits the dimensionality of the data, which determines the ability of spatial filtering techniques to resolve weak targets by separating them into a subspace distinct from the subspace that contains the noise. Here we propose to push back this limit by applying to the data a set of time shifts that increase the dimensionality and allow the source separation methods to synthesize time-domain finite impulse response (FIR) filters in addition to spatial filters. Time shifts were previously found to be useful to suppress environmental noise using reference sensors (de Cheveigné and Simon, 2007).

Concretely, the method produces a set $y_k(t)$, $k = 1, \dots, K$ of linear combinations of the time-delayed sensor channels:

$$y_k(t) = \sum_{j,d} a_{kjd} x_j(t-d) \quad (1)$$

where t is time, $d = 0, \dots, D-1$ is time shift, $x_j(t)$, $j = 1, \dots, J$ are channels of raw data, and a_{kjd} are weights. Components y_k are orthogonal and sorted in order of decreasing quality according to some criterion (see below). In particular, the first component is the *best linear combination* of time-shifted sensor channels, according to the chosen criterion. The second is the best linear combination

* Corresponding author at: Equipe Audition, ENS, 29 rue d'Ulm, F-75230 Paris, France. Tel.: +33 1 44322772.

E-mail address: Alain.de.Cheveigne@ens.fr.

orthogonal to the first, and so on. To obtain denoised data, a select subset of components $y_k(t)$, $j = 1, \dots, K'$ is projected back to sensor space:

$$\tilde{x}_j(t) = \sum_k b_{jk} x_j(t) \quad (2)$$

where $\tilde{x}_j(t)$, $j = 1, \dots, J$ are the cleaned sensor signals and b_{jk} are projection coefficients. In this paper we focus on the case $K' = 1$ where only the best component is retained. Additional components can be obtained by repeating the process after regressing out the best component (“deflation”).

The required weight parameters a_{kjd} are obtained using the DSS algorithm (Särelä, 2004; Särelä and Valpola, 2005). In brief, DSS involves the following steps: (a) orthogonalization by PCA of the data matrix (here augmented with time shifts), (b) normalization of the principal components, (c) application of a contrast function to emphasize target components and/or deemphasize noise, and (d) a final PCA to rotate the data so that the first component has the highest possible value of the contrast function, the second the highest value within the subspace orthogonal to the first, etc. Details of the algorithm are given in the next section, and related work is reviewed in Section 4.

2. Methods

2.1. Signal model

Sensor signals $\mathbf{X}(t) = [x_1(t), \dots, x_J(t)]^T$ include interesting brain activity and uninteresting noise activity:

$$\mathbf{X}(t) = \mathbf{X}_B(t) + \mathbf{X}_N(t). \quad (3)$$

The first term results from the superposition within each sensor of sources of interest within the brain, $\mathbf{X}_B(t) = \mathbf{A}_B \mathbf{S}_B(t)$, and the second results from uninteresting sources within the environment, sensors, and subject's body, $\mathbf{X}_N(t) = \mathbf{A}_N \mathbf{S}_N(t)$, where \mathbf{A}_B and \mathbf{A}_N are mixing matrices. Our aim is to attenuate $\mathbf{X}_N(t)$ so as to isolate $\mathbf{X}_B(t)$, thereby improving our observation of brain activity.

The sensor signals $\mathbf{X}(t)$ span a space of dimensionality J . Linear separation techniques such as ICA, beamforming or DSS assume that source and noise live in separate subspaces, the only difficulty being to find them. Here we are interested in the situation where target and interference components are partly colinear, and thus *not* separable using such techniques.

2.2. Contrast function

For clarity we will focus on the common situation where data are recorded in response to N presentations of the same stimulus, and we are interested in reproducible stimulus-evoked activity. The contrast function ϕ used in this paper is the average over trials:

$$\phi[\mathbf{X}(t)] = \left(\frac{1}{N}\right) \sum_{n=1}^N \mathbf{X}^n(t) \quad (4)$$

where $\mathbf{X}^n(t)$ designates data for trial n . Stimulus-evoked activity is reinforced by averaging, whereas stimulus-unrelated activity and noise are not, so the norm of $\phi[\mathbf{X}(t)]$ is greater for evoked activity than for noise. Other contrast functions can be used but they are not considered in this paper.

2.3. Algorithm

Given a data matrix of size $[J, T, N]$ (channel \times time \times trial), the following steps are performed:

1. Data are time-shifted over the range $0, \dots, D - 1$, the time-shifted data being treated like additional channels to form a matrix of size $[JD, T, N]$.
2. Data are reorganized by concatenating trials along the time dimension to obtain a matrix of size $[JD, TN]$. Each row corresponds to a particular channel/time shift, and each column to a particular time sample/trial.
3. PCA is performed, and the resulting component time series are normalized.
4. The data are reorganized again by trials to form a 3D matrix of size $[JD, T, N]$ (component \times time \times trial). The contrast function defined in Eq. (4) is applied, resulting in a matrix of size $[JD, T]$.
5. A second PCA is performed on this matrix.
6. The *rotation matrix* from the second PCA is applied to the whitened data from step 2 to obtain a matrix of components of size $[JD, T, N]$ (component \times time \times trial). Each component is of size $[T, N]$ (time \times trial).

Step 6 produces components that have two useful properties: (a) they are mutually orthogonal and (b) they are ordered by decreasing ratio of evoked power to total power. The first component is the *best* linear function of the data according to that criterion, and each subsequent component is the best within the subspace orthogonal to previous components. Eq. (1) defines a spatiotemporal filter: each component is a weighted sum of time-delayed channels. Intuitively it is easy to see why the algorithm maximizes evoked power: the first three steps create a data set that is “spherical” in the sense that all directions in multidimensional space have equal variance. Averaging over trial “tips the balance” by favoring certain directions that are selected by the second PCA. The evoked/raw power ratio is maximal along these directions.

The subspace spanned by the first few components contains the most reproducible brain activity. However, that subspace may also include time-shifted versions of brain activity, and thus all FIR-filtered versions of that activity, for filters of order up to D . Components produced by step 6 may thus differ from real brain activity by the effect of an arbitrary FIR filter of order D . This is not always a concern, as spectral filtering is commonly applied to improve quality of data. Nevertheless it may be desirable to minimize potential filtering effects and approximate unfiltered source activity. In that case an additional step may be performed:

7. Given a subset of components chosen for reproducibility, canonical correlation is used to find the linear combination most correlated to the raw data. Canonical correlation (Hotelling, 1936) is a technique that finds the linear combination of vectors within one set that best accounts for the variance in another set, and vice versa.

In the following we assume that only one component is retained (the most reproducible, or the most similar to unfiltered data within the chosen subspace). If stopped at step 6, the algorithm has one user-set parameter, D , if continued to step 7 an additional parameter is the number of components within the chosen subset.

2.4. Implementation

The algorithm was implemented in Matlab. A few details are worth mentioning. The size of the data matrices grows proportionally with D and that of the covariance matrices with D^2 , limiting the values of D that can be used. To save space, the time series may be divided into chunks and the covariance matrices calculated over these chunks and then summed. As in any calculation involving sums of squares, it is important to discount outliers. This is done by assigning them a zero weight in the calculations. PCA components with relative power below a threshold (e.g. 10^{-9}) are discarded in

step 3 to save memory and computation time and avoid numerical problems.

2.5. Evaluation data

The algorithm is evaluated first with synthetic test data, and then with data from several biomagnetic sensor arrays: small animal MEG, spinal cord magnetometer, and human MEG. Small animal MEG data came from a system equipped with a 9-channel array of magnetometers in a 8×8 mm array at 3 mm from the external surface of the dewar placed over the animal's head (Miyamoto et al., 2007, 2008). Three reference sensors sampled environmental noise, and an accelerometer sampled environmental vibration. Sensors and subject were placed within a magnetically shielded enclosure. Data were acquired at a sampling rate of 1000 Hz after high-pass filtering at 0.1 Hz and low-pass filtering at 200 Hz, and environmental and sensor noise attenuated using the TSPCA and SNS algorithms (de Cheveigné and Simon, 2007, 2008a). Data were recorded from guinea pig, gerbil and mouse in response to auditory stimuli presented diotically via flexible sound tubes ending at 1.5 cm from each ear and driven by Etymotics transducers placed outside the shielded chamber. Spinal cord data were recorded using a newly developed 120-channel vector magnetometer array placed within a dewar optimized for measurements from the spinal cord (Adachi et al., 2009). Data were filtered between 100 and 5000 Hz and recorded at a 10000 Hz sampling rate in response to 4000 repetitions of electric stimulation of a peripheral nerve. Human MEG data were acquired from a 160-channel system equipped with axial gradiometers (Kado et al., 1999). Data were sampled at 500 Hz after high-pass filtering at 0.1 Hz and low-pass filtering at 200 Hz, and environmental and sensor noise components were removed using the TSPCA and SNS algorithms. Data were recorded from human subjects in response to a set of auditory stimuli comprising clicks, short tones, noise bursts and tone ramps interleaved at random.

2.6. Evaluation statistics

The effect of denoising is quantified in terms of power (before vs. after denoising) and variance across stimulus repetitions (evoked/total power). The reliability of evoked responses is estimated using bootstrap resampling (Efron and Tibshirani, 1993). In brief: random sets of N trials are drawn with replacement, and the average over each set is calculated. This operation is repeated (typically 100 times) and the mean and standard deviation are calculated at each time sample within the epoch.

3. Results

3.1. Simulated data

3.1.1. Simulation 1

This simulation illustrates the limits of static methods such as DSS. The artificial data had dimensions 10 channels \times 1000 samples \times 100 trials. The “signal” was a single period of a sinusoid as illustrated in Fig. 1 (top), identical in all channels and repeated over all trials, and the “noise” consisted of 9 (resp. 10) independent Gaussian white noise sources mixed within each of the 10 channels via a 9×10 (resp. 10×10) mixing matrix with random coefficients. The SNR is defined as the ratio of signal and noise power summed over channels, transformed to decibels (dB) according to the formula $SNR = 10 \log_{10} P_s / P_n$.

Fig. 1 (bottom left) shows data for the “best” channel (the channel with the largest correlation with the clean signal) for $SNR = -40$ dB (RMS amplitude ratio 1/100, power ratio 1/10000). The blue line is the average over trials, and the gray band represents ± 2 SD of a

bootstrap resampling of that average. This plot is for 9 noise sources, the plot for 10 noise sources had a similar aspect (not shown). At this unfavorable SNR, the signal seems to be lost within the noise. The middle plot shows the first component of a DSS analysis (de Cheveigné and Simon, 2008b) for 9 noise sources. The recovered component is identical to the signal. In this case, signal and noise were separable within the 10-dimensional data space and DSS was successful in separating them. However, the right hand plot shows the same analysis for 10 noise sources, in which case the noise has the same dimensionality (10) as the data. Signal and noise are no longer linearly separable and the result of DSS is noisy (but nevertheless an improvement over no processing). This is the situation that we are interested in, where the dimensionality of the noise is too large to allow perfect linear separation from the signal.

3.1.2. Simulation 2

This simulation illustrates how TDDSS can be more effective than DSS. The signal was the same as in Simulation 1 (Fig. 1 top). The noise was produced from 5 independent noise sources mixed within each channel via a 5×10 random matrix, and the same 5 sources were shifted in time by 1 sample and added to each channel via another 5×10 random matrix. This noise has a similar covariance matrix as in the previous example, but there is now a serial correlation within and across channels.

Fig. 2 (top) shows the result of applying standard DSS at SNRs of -20 , -40 and -60 dB. DSS is quite successful at -20 dB and partly successful at -40 dB (as with 10 independent sources, Fig. 1 bottom right), but it performs poorly at -60 dB. The lower plot shows the result of applying TSDSS ($D = 1$). Now the signal is almost perfectly recovered even at -60 dB (1/1000 RMS ratio, 10^{-6} power ratio). The technique is thus successful in extracting a component with a power of 1 part per million (PPM) relative to the noise. For smaller SNRs (e.g. -80 dB) the algorithm may fail depending on the noise, and at yet smaller values it appears to fail systematically (not shown). Similar results were obtained for various other values of the D parameter and other noise correlation structures (not shown).

3.1.3. Simulation 3

This simulation tests TSDSS with noise correlation structures more complex than a simple delay. The signal was the same as in Simulations 1 and 2 (Fig. 1 top), but the noise was now produced by adding 50 sinusoids with frequencies multiples of 24.5 Hz (sampling rate was 1000 Hz), equal amplitudes, and phases that changed randomly from trial to trial, distributed over the 10 channels via a 50×10 random vector. Thus, within each channel the noise had a complex spectrum consisting of narrowband components modulated by an irregular spectral envelope that was different for each channel.

TSDSS was implemented with $D = 10$. Fig. 3 (top) shows that standard DSS is successful at -20 dB but fails at -40 and -60 dB. The reason for this failure is that the dimensionality of the noise (50) is much greater than that of the data (10), so noise and signal are not linearly separable. The bottom row shows that TSDSS is successful in each case.

These simulations illustrate situations where the TSDSS method can retrieve very weak targets with little distortion.

3.2. Real data

3.2.1. Small animal MEG

Data were recorded from a guinea pig in response to repeated tone pips of frequency 2 kHz, duration 30 ms, and 5 ms raised-cosine onset and offset ramps, presented diotically at a rate of 3 pips/s during a 1-h recording session, for a total of about 11000 trials of which 40% of were discarded as outliers (and to allow the data to fit in memory). These data have very poor SNR. After ini-

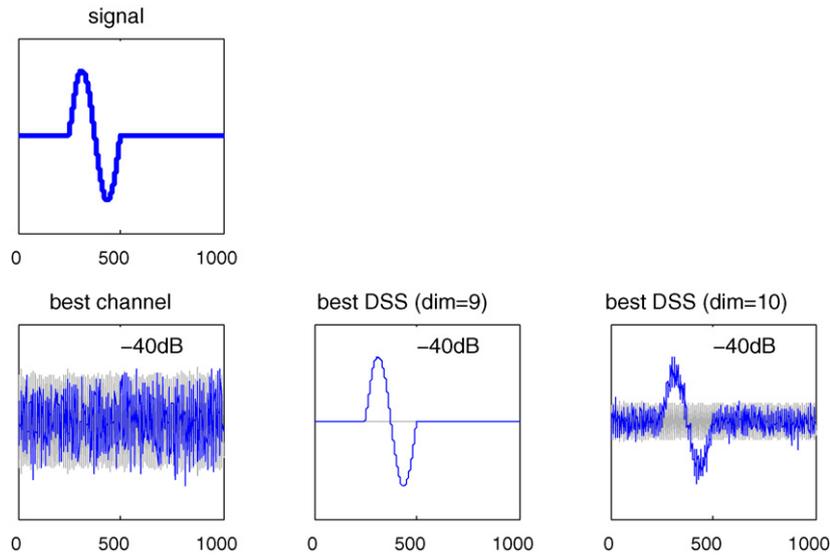


Fig. 1. Effect of noise dimensionality on the performance of DSS. Artificial data involve 10 channels, each a mixture of one signal source (identical in all channels and all trials) and either 9 or 10 independent Gaussian noise sources mixed with random coefficients. Overall SNR is -40 dB (1/100 in amplitude). All plots are based on 100 trials, blue is the mean over trials, gray indicates ± 2 standard deviations of a bootstrap resampling of the mean: (top) source waveform; (bottom left) waveform of the channel with best SNR; (middle) best DSS component for 9 noise sources; (right) best DSS component for 10 noise sources. DSS can extract the signal perfectly if target and source are linearly separable (middle), but not if they are not separable (right). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

tial denoising (see Section 2), the 9-channel data were subjected to three different processing schemes: PCA, DSS, and TSDSS with $D = 20$ as illustrated in Fig. 4 (top). Here PCA is used as a baseline to evaluate the improvement offered by DSS or TSDSS. Comparing the time course of the averages over trials (blue), the first principal component (left) appears to be noisier than the first DSS component (middle), itself noisier than the first TSDSS component (right), as confirmed by the bootstrap resampling (gray bands).

3.2.2. Spinal cord magnetometer

Data were recorded from a human subject in response to 4000 repetitions of an electrical stimulation of the median nerve of the left wrist. The 120-channel data were subjected to PCA, DSS, and

TSDSS ($D = 4$) as illustrated in Fig. 4 (bottom). The first TSDSS component (right) is less noisy than the first DSS component (middle) itself less noisy than the first principal component (left).

3.2.3. Human MEG

Data were recorded from a human subject in response to 50 repetitions of a 1 kHz tone pip of duration 100 ms presented diotically at a comfortable level. These were interleaved with various other stimuli within a 5-min recording session. The original data comprised 160 channels, but for this simulation only 7 channels were retained (to obtain data with low spatial dimensionality to better illustrate the behavior of the TSDSS algorithm). After initial denoising to remove environmental and sensor noise (see Section

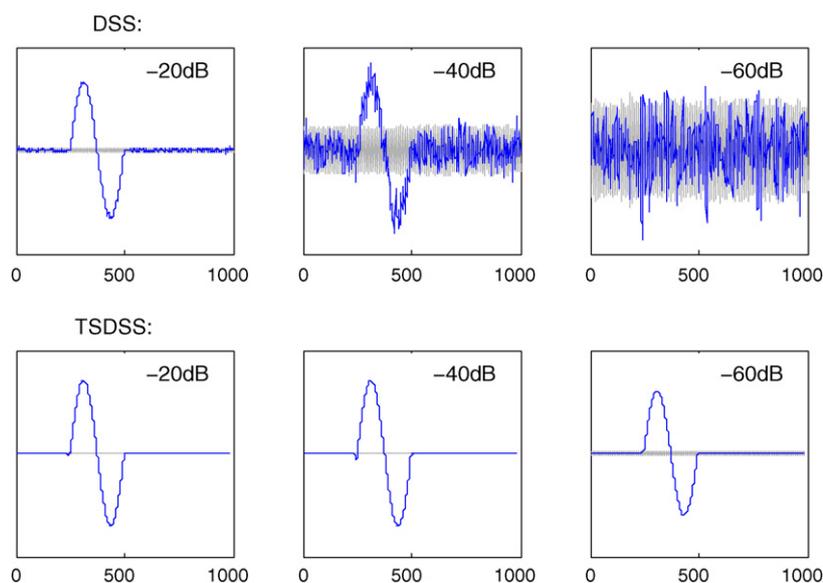


Fig. 2. Comparison of DSS and TSDSS on artificial data. The target is the same as in Fig. 1, the noise consists of 10 sources with serial correlation (see text) mixed with random coefficients. Overall SNR was -20 , -40 or -60 dB (1/10, 1/100 and 1/1000 in RMS amplitude). All plots are averages over 100 trials: (top) DSS algorithm; (bottom) TSDSS algorithm implemented with $D = 1$. The performance of DSS degrades as SNR decreases (top row), whereas that of TSDSS remains good for SNRs as low as -60 dB (1/1000 in amplitude, 1 PPM in power).

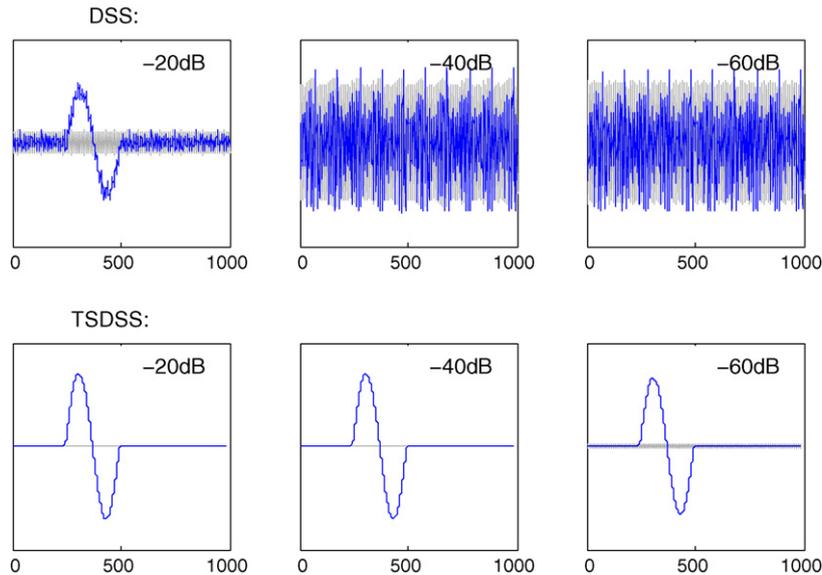


Fig. 3. The target is the same as in Figs. 1 and 2. The noise consists of 50 sinusoidal sources added with random amplitude and phase, different for each of the 100 trials: (top) DSS algorithm and (bottom) TSDSS algorithm implemented with $D = 10$. The performance of DSS degrades as SNR decreases (top row), whereas that of TSDSS remains good for SNRs as low as -60 dB (1/1000 in amplitude, 1 PPM in power).

2), data were processed by the DSS algorithm (Fig. 5 left), and by the TSDSS algorithm with $D = 30$ (Fig. 5 right). The blue line indicates the average over trials, and the gray band ± 2 SD of a bootstrap resampling. The TSDSS algorithm appears to yield a cleaner result, with a smoother average and narrower bootstrap interval. One might wonder whether the benefit of TSDSS is greater than

that obtained by simply smoothing the data with kernel of size $D = 30$. To answer this question, data were smoothed with such a kernel before applying DSS (middle). The result is less reproducible than with TSDSS, as reflected by the width of the bootstrap band (gray). With a full data set (150 channels rather than 7) the benefit of TSDSS was less impressive (not shown), presumably because that

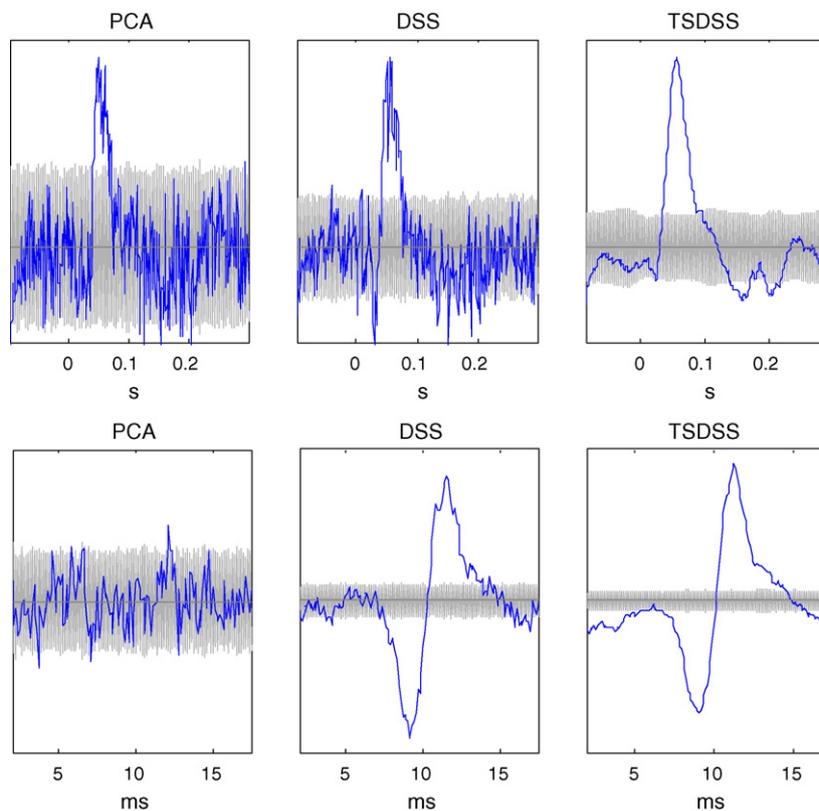


Fig. 4. Comparison of TSDSS and DSS on real data. (Top) Data from a 9-channel magnetometer for small animals. The fields originate from the auditory cortex of a guinea pig in response to 6400 repetitions of a 2 kHz tone pip of duration 30 ms: (left) first PCA component; (middle) first DSS component; (right) first TSDSS component ($D = 20$). (Bottom) Data from a 120-channel magnetometer for spinal cord evoked field measurements. The fields originate from the spinal cord of a human subject in response to 4000 repetitions of an electrical stimulation: (left) first PCA component; (middle) first DSS component; (right) first TSDSS component ($D = 4$).

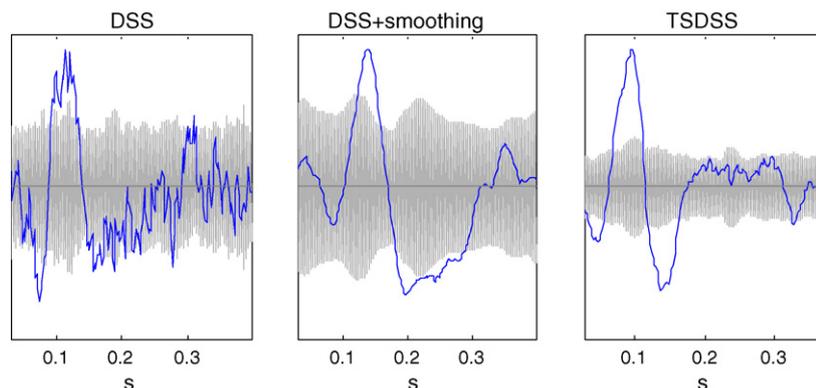


Fig. 5. Comparison of TSDSS and smoothing on real data. MEG data were recorded from a 160-channel whole-head axial gradiometer array from one human subject in response to 50 repetitions of a click-like stimulus. In order to illustrate the algorithm on data with reduced dimensionality (relative to that of cortical fields), 7 channels were arbitrarily selected for processing: (left) first DSS component; (right) first TSDSS component ($D = 30$); (middle) first DSS component after smoothing the data with a 30-sample square window. Note the wider bootstrap band for DSS with smoothing than for TSDSS.

number of channels was sufficient to represent the noise without delays.

4. Discussion

TSDSS performs by introducing time shifts that allow the algorithm to synthesize an optimal *multichannel FIR spatiotemporal filter*, rather than just a spatial filter. The benefit is clear on the artificial data, on real data it is more limited but nevertheless appreciable.

4.1. How it works

The principle of DSS is well explained by Särelä (2004); Särelä and Valpola (2005), and has been illustrated in the context of an evoked-response paradigm by de Cheveigné and Simon (2008b). In brief, the initial PCA and normalization steps of the DSS algorithm produce a data set that is “spherical” in the sense that it extends equally in every direction of the multidimensional signal space, and is thus free to rotate. The contrast function emphasizes certain directions of interest (here: directions of greatest repeatability), and the second PCA aligns them with the coordinate axes. In this way DSS finds the optimal linear combinations for the chosen criterion. The effectiveness of DSS depends on the dimensionality of the data, itself determined by the number of sensors. Intuitively, one might expect time shifts to be beneficial simply by increasing the dimensionality. Furthermore, the ability to synthesize multichannel FIR filters might give leverage in at least two ways. The first is by designing a spectral filter that emphasizes spectral features of the target relative to the noise (as in a Wiener filter), regardless of spatial structure. The second is by giving the algorithm more flexibility to suppress interfering sources on the basis of their spatial structure. As an example, suppose two noise sources such that the target is not collinear with either of them, but nevertheless falls within the hyperplane that they span. If either noise source were alone the target would be separable, but if both are active at the same time it is not. However, if the two noise sources have different spectral properties, TSDSS may be able to find a solution involving two spatial filters, each optimized for a different spectral region.

4.2. Caveats and cautions

The algorithm extracts weak sources by *cancelling* strong interference components. This delicate “balancing act” might go wrong in a variety of ways, a remark that was made previously for DSS (de Cheveigné and Simon, 2008b), but that applies even more to

TSDSS because of the larger number of parameters, and because TSDSS may be applied at lower SNRs. It is essential to understand what the algorithm is doing, and to corroborate its results by cross-validation.

The method involves a large number of free parameters (JDK' , where J , D and K' are the number of sensors, shifts, and retained components, respectively), and thus it is susceptible to over-fitting: even within a random data set there may exist linear combinations that appear “repeatable” to some degree. The investigator should parry this possibility by applying cross-validation tests. For example, the denoising stage may be included within the resampling loop of a bootstrap procedure (Efron and Tibshirani, 1993). Concretely: sets of trials are drawn with replacement, the denoising algorithm is applied to each set, and statistics (mean and standard deviation) are calculated over the population of denoised sets.

Another concern is spectral distortion. Suppose that the algorithm succeeds in suppressing all noise components. The SNR is now infinite, but the recovered data may nevertheless differ from true brain activity, as TSDSS components are weighted sums of sensor signals and their time-shifted versions. This can obviously affect the latency (within a range of size D), but also the shape of the response. To illustrate this point, Fig. 6 shows the result of a PCA on the time-shifted versions of the signal depicted in Fig. 1 (top left) for $D = 5$. The first principal component (PC) is a slightly-smoothed version of the signal, but none of the subsequent PCs resembles the signal. The subspace of time-shifted signals includes such distorted waveforms, and TSDSS components are potentially susceptible to similar distortion. One can argue that TSDSS components are optimized for repeatability, and thus any filtering will emphasize the “true” evoked response. Nevertheless the optional step 7 of the algorithm offers a means to minimize this distortion by rotating the subspace spanned by the selected components to maximize similarity to the undelayed data (at the expense of repeatability).

It is important to stress a potential pitfall of using TSDSS or other data-dependent algorithms such as DSS, ICA or beamforming. If the algorithm is applied separately to each condition of an experiment, differences in data-dependent processing might masquerade as effects. One way to avoid this is to interleave conditions to ensure that the same processing is applied to all conditions.

4.3. Relation with previous methods

The DSS method was proposed by Valpola and Pajunen (2000) and developed by Särelä (2004; Särelä and Valpola, 2005). It is related to the filtering by optimal projection (FOP) and common spatial pattern (CSP) algorithms (Boudet et al., 2007; Blankertz et

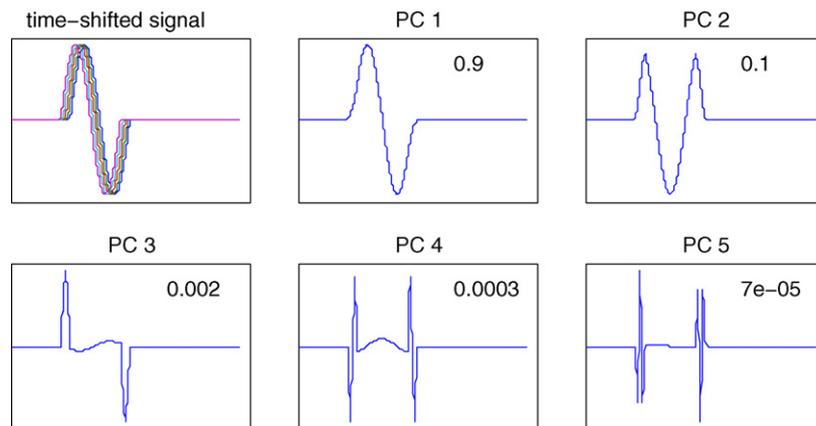


Fig. 6. Subspace spanned by a single-cycle sinusoidal waveform shifted by $[0, \dots, 4]$ samples (top left) as revealed by principal component analysis. Linear combinations of time-shifted signals can give rise to waveforms with shapes quite different from the original waveform. The number in each plot indicates the proportion of power represented by that component.

al., 2008; Koles et al., 1990) that have been widely used for classification in the brain computer interface (BCI) literature. These methods have earlier roots (Fukunaga, 1990; Parra et al., 2005). In de Cheveigné and Simon (2008b) we applied DSS (without time shifts) to denoising of stimulus-evoked data.

The idea of augmenting multichannel data with time shifts has been proposed in several contexts. The initial two steps of TSDSS correspond to the spatiotemporal whitening proposed by Beucker and Schlitt (1997). Time delays are involved in various other analysis techniques (Friman et al., 2002; Gruber et al., 2006; Sander et al., 2002; Woon and Lowe, 2004; Ziehe et al., 2000). TSDSS exploits the serial correlation structure of the data, that can also be captured in the frequency domain (Anemüller et al., 2003) or by multichannel linear prediction (e.g. Xie et al., 1992). Dornhege et al. (2006) have proposed to include time shifts in CSP to derive classification features in a brain computer interface application, a proposition closely related to ours. An appeal of DSS is that it is comparatively easy to understand, an advantage for non-specialized users.

4.4. Summary

TSDSS is effective in extracting extremely weak signals from noise. It capitalizes on the dimensionality of large multisensor arrays such as used in MEG, EEG or multielectrode electrophysiology, and on the spatiotemporal disparities between target and noise and/or between different noise components. In this respect, it pushes the limit of what can be done to retrieve signals at extremely unfavorable SNRs using purely linear techniques. Applied to data from human MEG, small animal MEG, and spinal cord biomagnetometry, TSDSS offered an improvement over DSS in the quality and reliability of extracted components as assessed by bootstrap resampling. A similar benefit was observed in simulations, for which the method succeeded over a wide range of conditions for which the best alternative algorithm that we know (DSS) failed. This suggests that there may be applications for which TSDSS is the only effective solution to extract exploitable data.

Acknowledgments

Thanks to Yoshiaki Adachi for providing the spinal cord data, and to Maria Chait for recording the human MEG data. Jonathan Simon, Israel Nelken and one anonymous reviewer provided useful comments on previous versions of this paper. This work was supported in part by BBSRC grant BB/H006958/1, by the Neuroinformatique 2009 program of CNRS, and by a collaboration grant

awarded by NTT. The small animal MEG system was provided by Kanazawa Institute of Technology.

References

- Adachi Y, Kawai J, Miyamoto M, Ogata H, Tomori M, Kawabata S, et al. A SQUID system for measurement of spinal cord evoked field of supine subjects. *IEEE Trans Applied Supercond* 2009;19:861–6.
- Ahonen AI, Hämäläinen MS, Ilmoniemi RJ, Kajola MJ, Knuutila ET, Simola JT, et al. Sampling theory for neuromagnetic detector arrays. *IEEE Trans Biomed Eng* 1993;40:859–68.
- Anemüller J, Sejnowski TJ, Makeig S. p. 47–52 Complex spectral-domain independent component analysis of electroencephalographic data. In: 4th international symposium on independent component analysis and blind signal separation; 2003.
- Beucker R, Schlitt HA. Temporal and spatial prewhitening of multichannel MEG data. Jülich Supercomputing Center technical report FZJ-ZAM-IB-9721; 1997.
- Boudet S, Peyrodie L, Gallois P, Vasseur C. Filtering by optimal projection and application to automatic artifact removal from EEG. *Signal Proc* 2007;87:1978–92.
- Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller K-R. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Process Mag* 2008;25:41–56.
- Cichocki A. Blind signal processing methods for analyzing multichannel brain signals. *Int J Bioelectromagn* 2004;6:1–18.
- de Cheveigné A, Simon JZ. Denoising based on time-shift PCA. *J Neurosci Methods* 2007;165:297–305.
- de Cheveigné A, Simon JZ. Sensor noise suppression. *J Neurosci Methods* 2008a;168:195–202.
- de Cheveigné A, Simon JZ. Denoising based on spatial filtering. *J Neurosci Methods* 2008b;171:331–9.
- Dornhege G, Blankertz B, Krauledat M, Losch F, Curio G, Müller K-R. Combined optimization of spatial and temporal filters for improving brain-computer interfacing. *IEEE Trans Biomedical Eng* 2006;11:2274–81.
- Efron B, Tibshirani RJ. Introduction to the bootstrap. Monographs on statistics and applied probability. Chapman and Hall/CRC; 1993.
- Friman O, Borga M, Lundberg P, Knutsson H. Exploratory fMRI analysis by autocorrelation maximization. *Neuroimage* 2002;16:454–64.
- Fukunaga K. Introduction to statistical pattern recognition. 2nd ed. Academic Press; 1990.
- Gruber P, Stadlthanner K, Böhm M, Theis FJ, Lang EW, Tomé AM, et al. Denoising using local projective subspace methods. *Neurocomputing* 2006;69:1485–501.
- Hotelling H. Relations between two sets of variants. *Biometrika* 1936;28:321–77.
- Kado H, Higuchi M, Shimogawara M, Haruta Y, Adachi Y, Kawai J, et al. Magnetoencephalogram systems developed at KIT. *IEEE Trans Appl Supercond* 1999;9:4057–62.
- Koles ZJ, Lazar MS, Zhou SZ. Spatial patterns underlying population differences in the background EEG. *Brain Topogr* 1990;2:275–84.
- Miyamoto M, Kawai J, Kawabata M, Shimozu T, Adachi Y, Uehara G, et al. Development of a biomagnetism measurement system for small animals. *Int Congr Series* 2007;1300:586–90.
- Miyamoto M, Kawai J, Adachi Y, Haruta Y, Komamura K, Uehara G. Development of an MCG/MEG system for small animals and its noise reduction method. *J Phys: Conf Series* 2008;97:012258.
- Parra LC, Spence CD, Gerson AD, Sajda P. Recipes for the linear analysis of EEG. *Neuroimage* 2005;28:326–41.
- Sander TH, Wubbeier G, Lueschow A, Curio G, Trahms L. Cardiac artifact subspace identification and elimination in cognitive MEG data using time-delayed decorrelation. *IEEE Trans Biomed Eng* 2002;49:345–54.
- Särelä J. Exploratory source separation in biomedical systems. Technical University of Helsinki, unpublished thesis; 2004.

- Särelä J, Valpola H. Denoising source separation. *J Mach Learn Res* 2005;6:233–72.
- Sekihara K, Nagarajan S, Poeppel D, Miyashita Y. Reconstructing spatio-temporal activities of neural sources from magnetoencephalographic data using a vector beamformer. *IEEE ICASSP* 2001;3:2021–4.
- Sekihara K, Hild KE, Nagarajan SS. A novel adaptive beamformer for MEG source reconstruction effective when large background brain activities exist. *IEEE Trans Biomed Eng* 2006;53:1755–64.
- Taulu S, Simola J, Kajola M. Applications of the signal space separation method. *IEEE Trans Biomed Eng* 2005;53:3359–72.
- Valpola H, Pajunen P. Fast algorithms for Bayesian independent component analysis. In: *Proceedings of the second international workshop on independent component analysis and blind signal separation (ICA 2000)*; 2000. p. 233–7.
- Vrba J, Robinson SE, McCubbin J. How many channels are needed for MEG? *Neurology, Neurophysiology and Neuroscience* 2004; 99:1–5, Available at: <http://www.neurojournal.com/article/view/313>. Date accessed: 12 November 2009.
- Woon WL, Lowe D. Can we learn anything from single-channel unaveraged MEG data? *Neural Comput Appl* 2004;13:360–8.
- Xie F, Le Cerf P, Van Compermolle D. Linear prediction analysis of multichannel speech recordings. In: *Proc. ProRisc IEEE Benelux workshop on circuits, systems and signal processing*, 1992; 277–82.
- Ziehe A, Müller K-R, Nolte G, Mackert BM, Curio G. Artifact reduction in magnetoneurography based on time-delayed second-order correlations. *IEEE Trans Biomed Eng* 2000;47:75–87.