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’Études 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 red2scartet9 - 1.343hl0ovrnate9Principmal
orthogonal to the first, and so on. To obtain denoised data, a select subset of components $y_j(t), j = 1, \ldots, K'$ is projected back to sensor space:

$$\hat{x}_j(t) = \sum_k b_{jk} x_j(t)$$

(2)

where $\hat{x}_j(t), j = 1, \ldots, J$ 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 $a_{ijd}$ are obtained using the DSS algorithm (Särrelä, 2004; Särrelä 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 de-emphasize noise, and (d) a final PCA to rotate the data so that the first component has the 9s9sc158a oscien8804rs
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 × 1000 samples × 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

\[ \text{SNR} = 10 \log_{10} \frac{P_s}{P_n} \]

Fig. 1 (bottom left) shows data for the “best” channel (the channel with the largest correlation with the clean signal) for SNR = −30 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 $-40 \text{ dB (1/100 in amplitude)}$. All plots are based on 100 trials, blue is the mean over trials, gray indicates $\pm 2$ 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.)

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 $-60 \text{ dB (1/10, 1/100 and 1/1000 in RMS amplitude)}$. 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 $-60 \text{ dB (1/1000 in amplitude, 1 PPM in power)}$. 

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. 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).

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 ($D = 4$).
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 Särélä (2004); Särélä 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 solve 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 carry 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 \(2.625 \text{ TDem-1753.3} (\text{signas9(1753.3(and)-157.3(thi.9(1753.3(tim-(shifend)-157.3(evensiosys...}

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.
represented by that component. 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. An 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 Yoshikatsu 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


Sekihara K, Hild KE, Nagarajan SS. A novel adaptive beamformer for MEG source