This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

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
Denoising based on time-shift PCA

Alain de Cheveigné a,b,∗, Jonathan Z. Simon c,d

a Laboratoire de Psychologie de la Perception, UMR 2929, CNRS and Université Paris Descartes, France
b Département d’Etudes Cognitives, Ecole Normale Supérieure, France
c Department of Electrical & Computer Engineering, University of Maryland at College Park, United States
d Department of Biology, University of Maryland at College Park, United States

Received 20 March 2007; received in revised form 1 June 2007; accepted 3 June 2007

Abstract

We present an algorithm for removing environmental noise from neurophysiological recordings such as magnetoencephalography (MEG). Noise fields measured by reference magnetometers are optimally filtered and subtracted from brain channels. The filters (one per reference/brain sensor pair) are obtained by delaying the reference signals, orthogonalizing them to obtain a basis, projecting the brain sensors onto the noise-derived basis, and removing the projections to obtain clean data. Simulations with synthetic data suggest that distortion of brain signals is minimal. The method surpasses previous methods by synthesizing, for each reference/brain sensor pair, a filter that compensates for convolutive mismatches between sensors. The method enhances the value of data recorded in health and scientific applications by suppressing harmful noise, and reduces the need for deleterious spatial or spectral filtering. It should be applicable to a wider range of physiological recording techniques, such as EEG, local field potentials, etc.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Magnetoencephalography (MEG); Electroencephalography (EEG); Noise reduction; Artifact removal; Artifact rejection; Regression; Principal component analysis

1. Introduction

Magnetoencephalography (MEG) measures magnetic fields produced by brain activity using sensors placed outside the skull. The fields to be measured are extremely small, several orders of magnitude below fields from unavoidable sources such as electric power lines, ventilators, elevators, or vehicles. Environmental noise is combated by a combination of magnetic and electromagnetic shielding, active noise field cancellation, the use of gradiometers, spectral and spatial filtering, averaging responses to repeated stimulus presentations, and various other signal-processing methods to reduce noise. MEG signals may also be contaminated by sensor noise arising in the quantum devices or associated electronics, and physiological noise from physiological activity other than of interest (a category that is study- or application-dependent). We focus on environmental noise, but our approach is complementary with techniques that deal with the other two types of noise.

Shielding, the primary method for noise reduction, involves placing the system and subject within a chamber lined with layers of aluminium and mu-metal. In a recent proposition, head and sensors are surrounded by a superconducting shield bathed in liquid helium (Volegov et al., 2004). A citive shielding has also been proposed (Platzek et al., 1999). However, the cost and bulk of shielding is an obstacle to widespread deployment of MEG in scientific and health applications (Okada et al., 2006; Papanicolaou et al., 2005). New applications such as brain–machine interfaces (BMI), and advances in MEG technology (e.g. the non-cryogenic system of Xia et al., 2006) make the perspective of systems without shield attractive. For an existing system better shielding may not be an option. A signal-processing alternative to reduce the level of noise is thus welcome.

A second measure is the use of gradiometer sensors, implemented in hardware or synthesized in software from magnetometer arrays (Baill et al., 2001; Vrba, 2000). There are nine components to the magnetic field gradient (three spatial derivatives of each of the three spatial components), but typical systems sample only a few: radial gradiometers measure the radial derivative of the radial component, and planar gradiometers one or two of its tangential gradients. Brain sources
produce fields with large gradients at nearby sensors, whereas most environmental sources are distant and produce a relatively homogenous field, that the gradiometer discounts. Gradiometers are also more sensitive to shallow than deep brain sources. This property may be useful in some cases, but there is no flexibility to tune or disable it without compromising environmental noise rejection. Sensor geometry could more easily be optimized for brain sensitivity if environmental noise were taken care of by other means.

A third approach is spectral filtering. Environmental noise is typically dominated by slowly varying fields from elevators, vehicles, etc., and by power line components at 60 Hz (or 50 Hz outside the US) and multiples, that may be attenuated by hardware filters before analog-to-digital conversion. A typical protocol involves a combination of a high-pass filter (e.g. 0.1 or 1 Hz) and a notch filter at 60 Hz, in addition to the mandatory anti-aliasing low-pass filter. Further filtering may be applied in software. Spectral filtering has two serious drawbacks. First, recordings are blind to eventual brain activity within the frequency bands that are rejected. Second, features of the time-course of activity are “smeared” on an interval equal to the duration of the impulse response, which is on the order of the inverse of the width of spectral transitions (e.g. about 1 s for a 1 Hz high-pass or 1 Hz wide notch filter). Temporal distortion is inconsistent with the common claim of “millisecond temporal resolution” for MEG, although its impact is hard to assess because impulse responses are rarely published. Data quality would be enhanced if spectral filtering could be avoided.

A fourth approach is spatial filtering. Linear combinations of sensor signals are formed to attenuate noise and/or enhance brain activity. Examples are synthetic gradiometers (already mentioned), the Laplacian (e.g. Kayser and Tenke, 2006), principal component analysis (PCA) (e.g. A. hissar et al., 2001; Kayser and Tenke, 2003, 2006; Spencer et al., 2001), independent component analysis (ICA) (e.g. Barbat et al., 2004; Makeig et al., 1996; Vigário et al., 1998), signal space projection (SSP) (Tesch et al., 1995), signal space separation (SSS) (Taulu et al., 2005), beamforming (e.g. Sekihara et al., 2001, 2006) and other linear techniques (Parra et al., 2005). Spatial filtering is useful to tease apart the activity of multiple sources within the brain. While it can also remove environmental noise, using it for that purpose constrains the options for brain source analysis (Nolte and Curio, 1999). Spatial filtering distorts the spatial signature of sources of interest, and forward models (required for source modeling) may need adjusting.

Finally, a very common procedure is to average responses over multiple repetitions of the stimulus. Stimulus-evoked brain activity adds constructively, while noise components tend to cancel each other out. Measurement of the steady-state response (SSR) in the frequency domain obeys the same principle. Drawbacks are that only repeatable “evoked” activity may be observed in this way, the signal-to-noise ratio (SNR) improvement is modest (it varies with the square root of repetitions), and the procedure is costly in experimental time. Effective denoising would allow cheaper experiments, and possibly useful recordings of single-trial activity.

To summarize, a wide range of noise-reduction tools is available. Together, they allow high quality measurements of brain activity, as evident from the MEG literature. Nevertheless, some have drawbacks that interfere with the observation of brain response morphology. For others, prior removal of environmental noise would allow them to be optimized for the purpose of brain source analysis.

Some MEG systems are equipped with reference sensors that measure environmental fields. Regression of brain sensor signals on the subspace spanned by reference sensor signals allows the contribution of environmental noise to be attenuated without the need for spectral filtering, or spatial filtering of the brain sensor array. Several methods have been proposed for that purpose (e.g. A. dachi et al., 2001; A. hmar and Simon, 2005; Volegov et al., 2004; V. rb and Robinson, 2001). Assuming that noise sources are distant and their fields homogenous, three sensors should suffice to capture the three spatial components of the noise field, regardless of the number of sources. However a larger number of reference sensors may be useful if field gradients differ between noise sources. Assuming instantaneous propagation at the relatively low frequencies of interest (Hämäläinen et al., 1993), the responses of two sensors to the same noise component differ only by a scalar factor. One should thus expect projection techniques to be highly effective. However, electromagnetic shielding is known to be frequency-dependent (Hämäläinen et al., 1993), and the electronics (flux lock loop, hardware filters) may introduce convolutive mismatch between channels, in which case scalar regression does not work well.

The method to be described extends these techniques by augmenting the array of reference signals by delayed versions of the same. The linear combination of delayed signals constitutes, in effect, a finite impulse response (FIR) filter that is applied to
be described in matrix notation as

\[ S_k(t) = \Lambda E(t) \]

\[ R(t) = B E(t) + R_N(t) \]  \hspace{1cm} (2)

where \( \Lambda = [a_k] \) and \( B = [b_j] \) are mixing matrices with \( a_k \) and \( b_j \) scalar and \( R_N(t) \) is reference sensor noise. Sensor noise is supposed negligible (the question is discussed further on). If the relation between noise and sensor signals is convolutive (filtering and/or delay) the same notation can be used supposing that each element \( a_k \) or \( b_j \) of the mixing matrices \( \Lambda \) or \( B \) represents an impulse response, and replacing multiplication by convolution in Eq. (2). For example:

\[ r_j(t) = (b_j * e_j)(t) \]  \hspace{1cm} (3)

where \( r_j(t) \) is the contribution of noise source \( i \) to sensor \( j \). The brain activity term \( S_k(t) \) in Eq. (1) presumably also reflects multiple sources within the brain, however we do not need to detail this dependency. To summarize the signal model, brain sensors and reference sensors pick up the same environmental noise sources, but the relation between noise and sensor signals may have convolutive properties that differ between brain and reference sensors.

### 2.2. Algorithm

The TSPCA algorithm is straightforward. First, the reference channels \( R(t) \) are time-shifted by a series of multiples of the sampling period, both positive and negative: \( R(t+n), n = -N/(2+1), \ldots, N/2 \). Second, the set of time-shifted reference signals is orthogonalized by applying PCA, to obtain a basis of \( \sqrt{N} \) orthogonal time-domain signals. Third, each brain sensor signal is projected onto this basis, and the projection removed. The result is the “clean” signal.

For brain sensor \( k \) the overall process can be described as

\[ \hat{s}_k(t) = s_k(t) - \sum_{j=1}^{N} \alpha_{kj}(n)r_j(t - n - N/2) \]  \hspace{1cm} (4)

where \( \hat{s}_k(t) \) is the cleaned signal and the \( \{\alpha_{kj}(n)\} \) emerge from the combination of orthogonalization and projection. Coefficient \( \alpha_{kj}(n) \) can be understood as the \( n \)th coefficient of an \( N \)-tap finite impulse response (FIR) filter applied to reference signal \( j \) before subtraction from brain signal \( k \). This filter is optimal, in a least-squares sense, to minimize the contribution of noise components to the brain sensor signal. Note that the brain sensor signals \( s_k(t) \) are not filtered, and thus there is no spectral distortion of brain activity \( s_k(t) \) (we return to this question later). Processing can be summarized in matrix notation as

\[ \hat{S} = IS - AR \]  \hspace{1cm} (5)

where \( I \) is the identity matrix, \( A = [a_k] \) the matrix of coefficients found by orthogonalization and projection, and \( R \) represents the set of time-shifted reference channel signals.

### 2.3. Implementation

The algorithm was implemented in Matlab. The number of taps is an arbitrary tradeoff: effectiveness, computational cost, and risk of overfitting, all increase with \( N \). The value \( N = 200 \) (shift range of \( \pm 200 \) ms for a 500 Hz sampling rate) was chosen for our simulations yielding 600 time-shifted reference channels. After PCA, components with variance (relative to the first) below an arbitrary threshold (10^{-6} in our simulations) were discarded to avoid numerical problems in the next steps. The algorithm can be applied to data blocks or files of arbitrary size: smaller blocks allow the algorithm to accommodate eventual fluctuations in reference/brain sensor relations, while larger blocks reduce the risk of overfitting. We typically used a block size of 10^3 samples (200 s), but we did not observe ill effects with larger or smaller sizes.

### 3. Results

We first evaluate the method with MEG data from one particular system to illustrate its effectiveness as a practical tool. Next we use synthetic data to quantify eventual side-effects. Later on we give more examples with data from other systems.

#### 3.1. MEG data

##### 3.1.1. Setup

Magnetic signals were recorded using a 160-channel, whole-head system with 157 axial gradiometer sensors that measure fields from the brain and 3 magnetometer reference sensors oriented along orthogonal directions (KIT, Kanazawa, Japan: Kado et al., 1999). The system is situated within a magnetically shielded room to reduce magnetic fields from the environment. Except where noted, dc and very low-frequency fields are removed by a high-pass filter in hardware at 1 Hz, line noise is suppressed by a notch filter at 60 Hz, and aliasing is prevented by a low-pass filter at 200 Hz (for 500 Hz sampling) or 400 Hz (for 1 kHz sampling).

##### 3.1.2. Empty machine

Fig. 1(a) (red) illustrates the power spectrum averaged over channels in normal conditions but with no subject within the system. It consists essentially of environmental power that has eluded magnetic shielding, cancellation by the gradiometers, and attenuation by hardware filters. The power spectrum is dominated by several sharp components at 120 Hz and beyond, several narrow modes at intermediate frequencies (10–120 Hz), and a diffuse distribution of low-frequency power below 10 Hz [expanded in Fig. 1(d), red].

Fig. 1(a) (blue) shows the power spectrum after applying our algorithm to the same data as in Fig. 1(a) (red). Ninety-eight percent of the variance has been discarded, leaving only 2% of residual noise power. Sharp high-frequency components are virtually eliminated, and mid-frequency peaks are greatly reduced. The dip near 60 Hz reflects the hardware notch filter, not noticeable in raw data because it coincides with the 60 Hz line power component (see below). The low-frequency region is
3.1.3. Brain activity

Fig. 1(c) illustrates data recorded with a subject performing an auditory task (Chait et al., 2005), before (red) and after (blue) denoising. Before denoising, the brain activity of the subject is hard to distinguish from environmental noise. After denoising the brain activity emerges more clearly. Assuming that brain activity and environmental noise are orthogonal, we can estimate the approximate power of the brain response by subtraction, and thus derive a rough estimate of the power ratio of the signal (defined in this context as activity other than environmental noise) to the estimated environmental noise (SNRE). Note that this definition of signal includes all activity other than environmental noise. After denoising, SNRE approaches 10 dB over the 0–20 Hz frequency range that includes many important components of brain activity, with a peak of about 20 dB just below 10 Hz (Fig. 1(d)). It should be stressed that these are “single-trial” data, without spatial filtering, spectral filtering other than in hardware, or averaging over epochs.

3.1.4. Recording without hardware filters

The previous responses were recorded with hardware high-pass and 60 Hz notch filters, as is standard in most MEG studies. As mentioned in Section 1, filtering distorts the observations and it would be useful to avoid it, if possible. Fig. 2 shows data recorded with high-pass and notch filters deactivated (in red). The waveform (Fig. 2(a)) is dominated by a 60 Hz component visible as a peak in the power spectrum (Fig. 2(b)), as well as slower fluctuations visible in Fig. 2(b) as a prominent peak at very low frequencies. After denoising, both are greatly reduced, by about 40 dB for the former and 35 dB for the latter. On average over the spectrum, the power has been reduced by about 99%. This suggests that, with adequate denoising, hardware filters could be omitted (however filters may still be required to avoid overloading of analog-to-digital converters by noise components if the resolution of the converters is insufficient).

3.1.5. Is the target distorted?

A obvious concern is whether denoising distorts brain activity. It was already mentioned that brain activity does not undergo spatial or spectral filtering (Eq. (4)) as long as reference channels do not pick up brain activity. Spurious correlations might conceivably appear by chance between brain and delayed-reference subspaces, in which case genuine brain components might be stripped together with the noise. However, given that brain and environmental activity are unrelated, the power of any such components should be small.

We tested this conclusion with synthetic data for which the target and noise were both known. In a first simulation we used a target consisting of wide-band Gaussian noise independent between channels. For “noise”, we used data recorded in the absence of a subject in the MEG machine, modified by subtracting the residual power (about 2%) leftover after denoising. This is very similar to real environmental noise, but with the
3.1.6. Is it reasonable to assume that reference sensors pick up no brain activity?

If the reference sensors pick up fields from the brain, it is possible that some brain components are removed together with the noise. For data from a real system this possibility cannot be ruled out completely, but two arguments suggest that leakage of brain activity into reference channels is too small to be of practical concern for our setup. First, if there were significant leakage, we would expect the power spectrum of the reference channel signals to differ according to whether a subject is present or not. Fig. 3(b) compares reference channel power spectra measured without (blue) and with a subject (red). The spectra differ in detail, as expected from different samples of ongoing environmental noise, but the difference does not follow the shape of the brain power (Fig. 3(a), green). Second, significant leakage should show up as brain-like characteristics within the residual (noise) signal removed by denoising, for example after averaging over many repetitions of a stimulus. No such characteristics were found (not shown).

Leakage of brain activity into reference channels appears to be negligible for our setup. However, if the reference sensors picked up more brain activity, or less environmental noise as might occur with better shielding, leakage could lead to significant subtraction of brain activity. This should be checked for before introducing the method to a particular machine.

3.1.7. Are delays useful?

With $N=1$ the method defaults to scalar regression (see Section 1). The amount of residual environmental noise as a function of $N$ is plotted in Fig. 4 (top, full line). As $N$ is increased from 1 to 200, the residual noise power drops from about 20% to about 2% while the power of the target (dashed line) is almost constant, with the result that the signal-to-environmental noise ratio (dB) becomes positive for about $N>8$. Multiple delays are obviously useful. Crucially, this shows that TSPCA is not indiscriminate as to how it removes power from a noisy signal. The power decrease affects only the noise (full line) but not significantly the target (dashed line).

The middle panels of Fig. 4 show the three reference-brain impulse responses for one particular brain channel for $N=200$, and the bottom panels of Fig. 4 show the amplitude and phase transfer functions of the third of these filters. The shapes, resulting from the automatic regression procedure, are not easily interpretable. Non-zero values of the impulse response at lags other than the origin reflect the fact that the corresponding lags contribute to reduce noise.

As described above, the power of the delays to better isolate and remove noise can arise simply from non-instantaneous mixing of noise across channels. Additionally, multiple delays can further aid noise reduction if some independent noise components are differentially spectrally filtered with respect to another (since the effect of summing and delaying noise channels can also create spectral filtering).
3.1.8. Are MEG data typically that noisy?

Our illustrations were based on data from one rather noisy MEG system, and one might wonder whether other systems would also benefit. Fig. 5 shows data from a variety of systems from different makers and installed in different locations (details in caption). For each system, the power spectrum of one channel is shown before (red) and after (blue) denoising. In each case the spectrum of the raw MEG data comprises low-frequency and line frequency harmonics that denoising removes. The benefit of TSPCA is not restricted to one particular system.

Reference channels were unavailable for two systems (MEG systems 2 and 3). To apply TSPCA nevertheless, we derived “synthetic reference channels” by applying ICA and selecting the three components with the largest proportion of dc and line noise. This appears to be effective, but it amounts to a form of spatial filtering and shares its potential drawbacks. Real reference sensors would be preferable. The green line in the plot for system 3 is the result of applying the SSS algorithm available with that system. TSPCA appears to be competitive with this implementation of that denoising method. The purpose of these examples is to show that TSPCA may be of use for a range of MEG systems. They should not be interpreted as reflecting the relative quality of systems or sites.

3.1.9. How does TSPCA compare with other methods?

Methods differ in their requirements and side-effects, and a level ground for comparison is hard to find. Easiest to compare are methods that use reference channels. Setting $N=1$, TSPCA is equivalent to scalar regression, a standard technique used in different forms (e.g. Volegov et al., 2004). From Fig. 4 (top) it is clear that TSPCA is superior to scalar regression for $N>1$.

We compared TSPCA with two other methods, CALM (Adachi et al., 2001) that is widely used with KIT/Yokogawa systems,
and Fast-LMS (Ahmar and Simon, 2005), a state-of-the-art LMS algorithm developed by our group. TSPCA surpasses both methods (Fig. 4, top). Any other signal-processing techniques can make use of reference channels (Haykin, 1991) but a comprehensive review is beyond the scope of this paper. Suffice to say that we are not aware of a method in widespread use with performance comparable to TSPCA.

Comparison with techniques that do not engage reference channels is of limited use because TSPCA can be used together with them. TSPCA alters neither spectral nor spatial characteristics, and it is fully compatible with noise reduction measures that precede it (passive or active shielding) or follow it (spectral or spatial filtering). Of interest is whether combining those methods with TSPCA offers an advantage over applying them alone. Data in Fig. 1 were recorded from gradiometers with hardware filters (1 Hz high-pass and 60 Hz notch); obviously applying TSPCA is an improvement over mere filtering, and Fig. 2 suggests that TSPCA might even replace such filters. Similar arguments can be made for spatial filtering, which is involved in a wide range of techniques (PCA, ICA, SSS, etc.).

Another standard approach to reduce noise (environmental and physiological) is to average responses over multiple repetitions of the same stimulus. Fig. 6 shows responses from an auditory study (Chait et al., 2005) averaged over 100 repetitions. Plotted in (a) is the root-mean-square field RMS averaged over channels before (red) and after (blue) denoising. The stimulus onset is at 0 ms and at about 100 ms appears the typical ‘M100’ onset response (Roberts et al., 2000). The field distribution over the sensor array shows a typical ‘auditory’ configuration (hemispherically antisymmetric pair of magnetic dipoles) that is visible in the raw data (b, left), but is much more clear in the denoised data (c, left). At about 200 ms post-onset, an additional peak is visible in the denoised data, with a similar ‘auditory’ configuration of opposite polarity. In the raw data, however, that peak is no more prominent than spurious peaks at other times (e.g. 400 ms), and the distribution in Fig. 6(b, right) is dominated by noise. TSPCA followed by averaging offers improvement over averaging alone.

3.1.10. Reference sensor noise

Reference sensor noise is typically small compared to the amplitude of the environmental fields, and unlikely to affect the outcome of the calculation of orthogonalization and projection matrices. However, reference sensor noise is injected into the denoised data via Eq. (5), and may contribute to the new noise floor remaining after TSPCA. Therefore it is especially important that the reference sensors exhibit minimal sensor noise. A nother way to reduce the impact of reference sensor noise is to increase the number of reference sensors beyond the number (usually 3) required to describe the environmental noise field, as redundant sensors allow sensor noise to be reduced (see de Cheveigné and Simon, submitted for publication).

4. Discussion

The TSPCA algorithm has the following useful features:

- It is effective in removing environmental noise: in our simulations the single-trial SNR improved from −10 dB to about +10 dB overall.
- It does not involve spectral or spatial filtering, and thus does not distort brain activity.
- It is relatively efficient and easy to implement, and should be suitable for a real-time implementation in BMI applications.
- Once it has been validated for a system, it is suitable as a systematic unsupervised data preprocessing tool. It does not require tuning, calibration, component selection, or other expert intervention.
- It is applicable to recordings other than MEG. So far only EEG has been tested, but it is expected that the technique might benefit electrophysiology in general.
- It is complementary (and compatible) with other methods of noise reduction and source analysis.

The method does not address other sources of noise such as sensor noise or unwanted physiological activity such as heartbeat, eyeblinks, muscle activity, brain activity other than of interest, etc. Other noise-reduction or data analysis techniques are available for that purpose, with which TSPCA is complementary. Note that, if an independent measurement of a physiological
artifact is available, TSPCA may be used to optimize the rejection of that artifact, and that it is possible to include non-linear transforms in addition to delays, for example to compensate for eventual sensor non-linearities.

Effective denoising can replace spectral and spatial filtering, but hardware high-pass or notch filters may nevertheless be necessary to preserve dynamic range. Eq. (4) suggests, and simulations confirm, that the method does not appreciably distort