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

MEG (MEG)

sensor noise

physiological noise

gradiometer

A. de Cheveigné

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 (A. de Cheveigné).

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 (V olegov et al., 2004). Active 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 (Baillet 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
ponent analysis (ICA) (e.g. Barbati et al., 2004; Makeig et al., 2001; Tenke, 2003, 2006; Spencer et al., 2001), independent component analysis (PCA) (e.g. Ahissar et al., 2001; Kayser 1996; Vigário et al., 1998), signal space projection (SSP) (Tesche techniques (Parra et al., 2005). Spatial filtering is useful to tease beamforming (e.g. Sekihara et al., 2001, 2006) and other linear mentioned), the Laplacian (e.g. Kayser and Tenke, 2006), principal component analysis (PCA) (e.g. Ahissar et al., 2001; Kayser 1996; Vigário et al., 1998), signal space projection (SSP) (Tesche techniques (Parra et al., 2005). Spatial filtering is useful to tease beamforming (e.g. Sekihara et al., 2001, 2006) and other linear mentioned), the Laplacian (e.g. Kayser and Tenke, 2006), principal component analysis (PCA) (e.g. Ahissar et al., 2001; Kayser 1996; Vigário et al., 1998), signal space projection (SSP) (Tesche 1996; Adachi et al., 2001; Ahmar and Simon, 2005; V olegov et al., 2004; Vrba 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, whereas the gradiometer discounts. Gradiometers produce fields with large gradients at nearby sensors, whereas gradiometers placed over the brain and reference sensors may be useful if field gradients differ between sensors placed far from the brain and oriented orthogonally to the sensors. Several methods have been proposed for that purpose (e.g. about 1 s for a 1 Hz high-pass or 1-Hz wide-notch filter). A typical protocol involves a combination of a high-pass filter (or 50 Hz outside the US) and multiples, that may be attenuated by hardware filters before analog-to-digital conversion. Further filtering may be applied in software. Spectral filtering has two seri-ous drawbacks. First, recordings are blind to eventual brain sensitivity if environmental noise were taken care of by other means.

2. Methods

2.1. Signal model

\[
S(t) = S(t) + S(t) + S(t) + S(t) + S(t) + S(t)
\]

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. Denoting vectors with bold-faced letters, the linear combination of delayed signals constitutes, in this way, the signal-to-noise ratio (SNR) improvement is avoided.

Data quality would be enhanced if spectral filtering could be applied in software. Spectral filtering has two seri-ous drawbacks. First, recordings are blind to eventual brain sensitivity if environmental noise were taken care of by other means.

\[
A = \text{spectral filtering. } E
\]

\[
\begin{align*}
\{ \text{w} & \} \\
\{ \text{w} & \} \\
\{ \text{w} & \}
\end{align*}
\]

If the relation between each noise source and each sensor were scalar (no filtering or delay), the dependency could

\[
A = \text{spatial filtering. } L
\]

\[
\begin{align*}
\{ \text{w} & \} \\
\{ \text{w} & \} \\
\{ \text{w} & \}
\end{align*}
\]

For others, prior removal of environmen-tal noise may be applied in software. Spectral filtering has two seri-ous drawbacks. First, recordings are blind to eventual brain sensitivity if environmental noise were taken care of by other means.

\[
A = \text{spatial filtering. } L
\]

\[
\begin{align*}
\{ \text{w} & \} \\
\{ \text{w} & \} \\
\{ \text{w} & \}
\end{align*}
\]

2. Methods

2.1. Signal model

\[
S(t) = S(t) + S(t) + S(t) + S(t) + S(t) + S(t)
\]

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. Denoting vectors with bold-faced letters, the linear combination of delayed signals constitutes, in this way, the signal-to-noise ratio (SNR) improvement is avoided.

Data quality would be enhanced if spectral filtering could be applied in software. Spectral filtering has two seri-ous drawbacks. First, recordings are blind to eventual brain sensitivity if environmental noise were taken care of by other means.

\[
A = \text{spectral filtering. } E
\]

\[
\begin{align*}
\{ \text{w} & \} \\
\{ \text{w} & \} \\
\{ \text{w} & \}
\end{align*}
\]

If the relation between each noise source and each sensor were scalar (no filtering or delay), the dependency could

\[
A = \text{spatial filtering. } L
\]

\[
\begin{align*}
\{ \text{w} & \} \\
\{ \text{w} & \} \\
\{ \text{w} & \}
\end{align*}
\]

For others, prior removal of environmen-tal noise may be applied in software. Spectral filtering has two seri-ous drawbacks. First, recordings are blind to eventual brain sensitivity if environmental noise were taken care of by other means.

\[
A = \text{spatial filtering. } L
\]

\[
\begin{align*}
\{ \text{w} & \} \\
\{ \text{w} & \} \\
\{ \text{w} & \}
\end{align*}
\]

2. Methods

2.1. Signal model

\[
S(t) = S(t) + S(t) + S(t) + S(t) + S(t) + S(t)
\]

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. Denoting vectors with bold-faced letters, the linear combination of delayed signals constitutes, in this way, the signal-to-noise ratio (SNR) improvement is avoided.

Data quality would be enhanced if spectral filtering could be applied in software. Spectral filtering has two seri-ous drawbacks. First, recordings are blind to eventual brain sensitivity if environmental noise were taken care of by other means.

\[
A = \text{spectral filtering. } E
\]

\[
\begin{align*}
\{ \text{w} & \} \\
\{ \text{w} & \} \\
\{ \text{w} & \}
\end{align*}
\]

If the relation between each noise source and each sensor were scalar (no filtering or delay), the dependency could

\[
A = \text{spatial filtering. } L
\]

\[
\begin{align*}
\{ \text{w} & \} \\
\{ \text{w} & \} \\
\{ \text{w} & \}
\end{align*}
\]

For others, prior removal of environmen-tal noise may be applied in software. Spectral filtering has two seri-ous drawbacks. First, recordings are blind to eventual brain sensitivity if environmental noise were taken care of by other means.

\[
A = \text{spatial filtering. } L
\]

\[
\begin{align*}
\{ \text{w} & \} \\
\{ \text{w} & \} \\
\{ \text{w} & \}
\end{align*}
\]
2.2. Algorithm

\[ S(t) = AE(t) \]
\[ R(t) = BE(t) + R(t) \]

\[ w \frac{A = a_{lj}}{b_{jl}} B = b_{jl} \quad R(t) \quad s_j \quad \text{(2)} \]

\[ \text{w} \quad \frac{1}{N} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ 2.3. Implementation \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

2.3. Implementation

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

2.3. Implementation

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

2.3. Implementation

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

2.3. Implementation

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

2.3. Implementation

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

2.3. Implementation

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

2.3. Implementation

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

2.3. Implementation

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

2.3. Implementation

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

2.3. Implementation

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

2.3. Implementation

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

2.3. Implementation

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

2.3. Implementation

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]

\[ \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \quad \text{w} \]
3.1.3. Brain activity

\( \text{Fig. 1:} \) MEG responses before and after denoising. (a) Power spectrum recorded from an empty machine averaged over all channels, before (red) and after (blue) denoising. (b) Same, with an enlarged abscissa. (c) Same, in the presence of a subject. (d) Estimated signal-to-environmental noise ratio (SNRE) of brain fields before (black) and after (green) denoising. The estimate of SNRE before denoising was made by comparing power recorded with and without a subject in the MEG machine. The estimate of SNRE after denoising was made by comparing power after denoising to power before denoising. Both estimates are rough approximations.

3.1.4. Recording without hardware filters

Fig. 2 shows data recorded with high-pass and 60 Hz 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?

An 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...
frequencies where brain activity is intense suggest that leakage is weak. Lack of a systematic difference between reference power spectra (red, blue) at recorded in the absence of a subject. Red: same in the presence of a subject. The activity into reference channels. Blue: power spectrum of reference channels same after denoising. (b) The bottom plot illustrates estimating leakage of brain synthetic "brain activity". Red: same after addition of synthetic "noise". Blue: activity (blue) plotted over a 0–50 Hz range. Denoising suppresses the brain power (Fig. 3(a), green). Second, significant leakage might occur if the number of data samples were small relative to the brain power (Fig. 3(a), green). Third, the residual noise power drops from about 20% to about 2% while the power of the target (dashed line) is almost constant, to 200, the residual noise power drops from about 20% to about 2% while the power of the target (dashed line) is almost constant.

3.1.6. Is it reasonable to assume that reference sensors pick up no brain activity? Our target is the result of a denoising process and thus conceivably less susceptible to distortion than "real" brain activity, but noise components, but the target itself is not seriously distorted: the noise-contaminated activity (red), and the denoised activity (blue) plotted over a 0–50 Hz range. Denoising suppresses the noise. For data from a real system this possibility cannot be ignored. 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 ignored.

3.1.7. Are delays useful? As described above, the power of the delays to better isolate the brain activity (Fig. 3(b)) and remove noise can arise simply from non-instantaneous mixing of noise across channels. Additionally, multiple delays can contribute to reduce noise.

A second simulation used as a target data recorded from the brain while the subject is performing an auditory task, denoised by application of TSPCA. This is our best approximation, in terms of amplitude and spectral content, of brain activity as measured in sensor space (real brain activity being obviously inaccessible). We would expect the power spectrum of the reference channel 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 environmental noise, but the difference does not follow the shape of the result that the signal-to-environmental noise ratio (dB) was varied from 1 to 200 (not shown).

The amount of residual environmental noise as a function of amplitude and spectral content, of brain activity as measured in sensor space (real brain activity being obviously inaccessible). Overfitting produced a magnetic field synchronized to the stimulus, or if reference sensors picked up appreciable brain activity. Overfitting might occur if the number of data samples were small relative to the number of free parameters in the model (600 for $w = 200$, and the number of time samples $N = 200$).

3.1.7. Are delays useful? As described above, the power of the delays to better isolate the brain activity (Fig. 3(b)) and remove noise can arise simply from non-instantaneous mixing of noise across channels. Additionally, multiple delays can contribute to reduce noise.

A MEG signal was varied 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, to 200, the residual noise power drops from about 20% to about 2% while the power of the target (dashed line) is almost constant. The amount of residual environmental noise as a function of amplitude and spectral content, of brain activity as measured in sensor space (real brain activity being obviously inaccessible). Overfitting produced a magnetic field synchronized to the stimulus, or if reference sensors picked up appreciable brain activity. Overfitting might occur if the number of data samples were small relative to the number of free parameters in the model (600 for $w = 200$, and the number of time samples $N = 200$).

3.1.7. Are delays useful? As described above, the power of the delays to better isolate the brain activity (Fig. 3(b)) and remove noise can arise simply from non-instantaneous mixing of noise across channels. Additionally, multiple delays can contribute to reduce noise.
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 a single 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

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. Volovskii 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,
Fig. 6. MEG responses of one subject to an auditory stimulus, averaged over 100 repetitions (Chait et al., 2005). (a) Time-course of RMS over all channels before (red) and after (blue) denoising. (b) Topography of field over subject’s head before denoising at ∼ 100 and ∼ 200 ms post-stimulus onset. (c) Topographies after denoising.

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.).

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.

Another 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 Chevigné 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 heart-beat, 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...
that 98% of noise variance was removed, in particular within
Tests with data recorded from an empty MEG system found
for any mismatch between data and reference sensor channels.
filters that are optimal (in a least-squares sense) to compensate
the time-shifted reference signals. This effectively synthesizes
noise. Sensor channels are projected on a subspace spanned by

5. Conclusions

Acknowledgments

References


