Denoising based on time-shift PCA

Alain de Cheveigné\textsuperscript{a,b,∗}, Jonathan Z. Simon\textsuperscript{c,d}

\textsuperscript{a} Laboratoire de Psychologie de la Perception, UMR 2929, CNRS and Université Paris Descartes, France
\textsuperscript{b} Département d’Études Cognitives, Ecole Normale Supérieure, France
\textsuperscript{c} Department of Electrical & Computer Engineering, University of Maryland at College Park, United States
\textsuperscript{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). 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
be described in matrix notation as

$$S_k(t) = AE(t)$$
$$R(t) = BE(t) + R_k(t)$$

(2)

where $A = [a_{kj}]$ and $B = [b_{jl}]$ are mixing matrices with $a_{kj}$ and $b_{jl}$ scalar and $R_k(t)$ 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 $a_{kj}$ or $b_{jl}$ of the mixing matrices $A$ or $B$ represents an impulse response, and replacing multiplication by convolution in Eq. (2). For example:

$$r_{jl}(t) = (b_{jl} \ast e_j)(t)$$

(3)

where $r_{jl}(t)$ is the contribution of noise source $l$ 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 is convolutive (filtering and attenuation by hardware filters). The power spectrum is dominated by several sharp components at 120 Hz and beyond, and attenuation by hardware filters. The power spectrum is eluded magnetic shielding, cancellation by the gradiometers, and attenuation by hardware filters. 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 $10^5$ 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

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.

3.1.3. 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?

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
nice property that denoising removes it completely so that target distortion may be observed in isolation. Target and noise were added in sensor space to produce synthetic “noise-contaminated data” that were then processed by the TSPCA algorithm to obtain “denoised data”. After denoising, target power was reduced (uniformly over the spectrum) by less than 1 dB as $N$ was varied from 1 to 200 (not shown).

A second simulation used as a target data recorded from the MEG with a subject 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). Fig. 3(a) shows the power spectrum of the brain activity (green, thick), the noise-contaminated activity (red), and the denoised activity (blue) plotted over a 0–50 Hz range. Denoising suppresses noise components, but the target itself is not seriously distorted: comparing the green and blue plots, the differences are small. Our target is the result of a denoising process and thus conceivably less susceptible to distortion than “real” brain activity, but it is our best approximation in the absence of direct access to brain source activity.

Taken together, these arguments suggest that it is safe to assume that brain activity is not significantly distorted by TSPCA. Note that the assumption of uncorrelated brain and envi-
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 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,
303

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. 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. 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 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 SNRE 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
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 brain activity. This implies that forward models do not need to be modified, and the method can be used together with techniques such as source modeling, PCA, ICA, SSA, etc. (Ahissar et al., 2001; Baillet et al., 2001; Makeig et al., 1996; Parra et al., 2005). Indeed, removing a major source of noise may help make those techniques more effective.

Reference sensors must be available, although we saw that TSPCA can make use of a "synthetic reference". Regression on a synthetic reference amounts to a form of spatial filtering, and real reference sensors should be preferred if available. References should not be sensitive to physiological fields of interest. This should be verified when the method is applied to a new system, either directly with phantom sources, or indirectly by looking for traces of brain activity in the reference signals.

Our method extends previous methods that perform regression on reference sensor signals (Adachi et al., 2001; Volegov et al., 2004; Vrba and Robinson, 2001). It is superior to those methods in that it augments the reference signals with time-shifted versions of the same, thus allowing the synthesis of filters that compensate for eventual latency or filtering mismatches. In this respect it resembles frequency-domain regression (e.g. Vrba, 2000; Woestenburg et al., 1983). It can be understood loosely as a way to enhance the effectiveness of regression by compensating for convolutional mismatch. It should be applicable to other sources of artifact for which a brain-independent measurement is available, such as heartbeat or eye movements, and to other measurement techniques such as EEG.

This new MEG denoising technique is related to dynamic PCA used in process control (Ku et al., 1995), singular spectrum analysis (SSA) used in geophysics (Allen and Smith, 1997; Ghil et al., 2002; Vautard and Ghil, 1989), the delayed coordinate methods of Gruber et al. (2006), or the delayed correlation ICA methods of Ziehe et al. (2000) or Sander et al. (2002). All of these techniques involve augmenting a set of signals with delayed versions. To the best of our knowledge this is the first application of such ideas to MEG or EEG noise suppression (see however He et al., 2004).

5. Conclusions

The TSPCA method is effective for denoising MEG signals on the basis of reference channels that pick up environmental noise. Sensor channels are projected on a subspace spanned by the time-shifted reference signals. This effectively synthesizes filters that are optimal (in a least-squares sense) to compensate for any mismatch between data and reference sensor channels. Tests with data recorded from an empty MEG system found that 98% of noise variance was removed, in particular within frequency bands important for the study of brain responses.

While recording from a subject during an auditory task, estimated single-trial signal-to-noise ratios approaching 10 dB were obtained across the low-frequency band (0–20 Hz), with a peak of 20 dB at about 10 Hz. The method is of considerable practical interest, as it may allow MEG systems to be designed more cheaply, to be deployed in less controlled (especially clinical) environments, and require less time per experiment. It may be of use to improve the quality of information about the brain that is gathered by this brain imaging technique, as well as other recording techniques sensitive to noise.

Acknowledgments

Author JZS was supported by NIH-NIBIB grant 1-R01-EB004750–01 (as part of the NSF/NIH Collaborative Research in Computational Neuroscience Program). We thank Maria Chait for providing the data used to develop and evaluate the method, and for stimulating and critical discussions. Juanjuan Xiang and Nayef Ahmar offered technical help and useful discussions. Thanks to Israel Nelken for insight, to Sylvain Baillet and John Mosher for critical comments, to Jeff Walker for excellent technical support, and to Kaoru Amano, Minae Okada, Rhodri Cusack, Yasuhiro Haruta and Denis Schwartz for providing additional data examples. This work was initiated while author AdC was visiting researcher at Shihab Shamma’s Neural Systems Lab, Institute of Systems Research, University of Maryland. A previous version of this paper was submitted to journals NeuroImage in March 2006 and Journal of Neurophysiology in September 2006, and the reviewers of those submissions are acknowledged for their constructive criticism.

References

de Cheveigné A, Simon JZ. Sensor noise suppression; submitted for publication.


