MEG SIGNAL DENOISING BASED ON TIME-SHIFT PCA

Alain de Cheveigné*, Jonathan Le Roux* † and Jonathan Z. Simon ‡

*CNRS, Université Paris 5 and Ecole Normale Supérieure, Alain.de.Cheveigne@ens.fr
†Information Science and Technology, The University of Tokyo, leroux@hil.t.u-tokyo.ac.jp
‡Electrical & Computer Engineering/Biology, University of Maryland, jzsimon@isr.umd.edu

ABSTRACT

We present a method for removing environmental noise from physiological recordings such as Magnetoencephalography (MEG) for which noise-sensitive reference channels are available. Sensor signals are projected on a subspace spanned by the reference channels augmented by time-shifted and/or nonlinearly transformed versions of the same, and the projections are removed to obtain "clean" sensor signals. The method compensates for scalar, convolutional or non-linear mismatches between sensor and reference channels by synthesizing, for each reference/sensor pair, a filter that is optimal in a least-squares sense for removal of the artifact. The method was tested with synthetic and real MEG data, typically removing up to 98% of noise variance. It offers an alternative to bulky and costly magnetic shielding (multiple layers of aluminium and mu-metal) for present scientific and medical applications and future developments such as brain-machine interfaces (BMI).

Index Terms— Magnetoencephalography, Bioelectric potentials, Biomedical imaging, Interference suppression

1. INTRODUCTION

In magnetoencephalography (MEG), superconducting quantum interference device (SQUID) sensors placed outside the skull measure magnetic fields produced by brain activity [2]. Brain fields are extremely small, several orders of magnitude below environmental noise produced by sources such as electric power lines and mechanical elevators. Despite measures such as magnetic and electromagnetic shielding (multiple layers of aluminium and mu-metal), active noise field cancellation, and the use of gradiometers which are more sensitive to the inhomogenous field produced by proximal sources than the homogenous field of more distant noise sources [3, 4], recorded signals may be dominated by noise. Shielding, in particular, is expensive and bulky, and this is an obstacle to the deployment of MEG for scientific and medical applications, or the development of practical brain-machine interfaces (BMI).

Several noise-reduction techniques use reference sensors that respond primarily to noise sources to estimate their contribution to brain sensors, and remove that contribution. Our method follows the same spirit but with an original twist: sensor channels are augmented by various transforms (delays and non-linearity) that allow it to handle convolutive and/or nonlinear mismatches between sensors. The examples that we give to illustrate the method are from an MEG machine equipped with three magnetometer sensors sensitive to the homogenous field of distant sources. The method is however of use for other MEG configurations, and for other sensitive physiological recording techniques such as EEG, local field potentials, single-units, etc. for which environmental noise is a problem.

The method is a generalization of Principal Component Analysis (PCA). PCA is a linear transformation that "rotates" a set of data of dimension K, expressing each as a sum of K one-dimensional components ("principal components") that are (a) mutually orthogonal to each other, and (b) ordered in terms of variance from large to small. The total variance (or power) is conserved. Components are ordered with decreasing variance, and therefore as much variance as possible is "packed" into the first components. If the data set is of lower dimensionality than the space, i.e. it fits within a "hyperplane", later components may be discarded without loss. PCA is thus useful for dimensionality reduction. PCA is often used as an ingredient in denoising algorithms [5, 6, 7, 8, 9]. Here we use PCA in combination with subspace projection to synthesize, for each reference/target channel pair, a filter that maximizes the proportion of noise that can be suppressed. The goal is to provide a simple and effective means for reducing the impact of environmental noise on data recorded from MEG or other noise-sensitive techniques.

2. SIGNAL MODEL

Two sets of signals are observed: data sensor signals $\mathbf{s}(t) = [s_1(t), \cdots, s_K(t)]^\top$ and noise reference signals $\mathbf{r}(t) = [r_1(t), \cdots, r_J(t)]^\top$. The data sensor signals reflect a combination of brain activity and environmental noise:

$$\mathbf{s}(t) = \mathbf{s}_B(t) + \mathbf{s}_E(t) \tag{1}$$

whereas the J reference sensors reflect only noise. We approximate this situation using three signal models of increas-

ing complexity. In a first model, the multiple sources of environmental noise contribute to both sets of sensors via *scalar* mixing matrices:

$$\mathbf{s}_E(t) = \mathbf{An}(t) \qquad (2)$$
$$\mathbf{r}(t) = \mathbf{Bn}(t)$$

where $\mathbf{A} = [a_{kl}]$ and $\mathbf{B} = [b_{jl}]$ are the mixing matrices and $\mathbf{n}(t) = [n_1(t), \cdots, n_L(t)]^\top$ represents the L noise sources. The brain term, $\mathbf{s}_B(t)$, also depends on multiple sources within the brain but the details of this dependency do not interest us here. Scalar mixing is not an unreasonable assumption, given that the propagation of magnetic fields is practically instantaneous. A second model assumes a similar dependency of sensor signals on noise sources, but via *convolutional* mixing matrices. Assuming convolution instead of multiplication in Eq. 2, each element a_{kl} or b_{jl} of the mixing matrices $\mathbf{A}$ or $\mathbf{B}$ now represents an *impulse response*, e.g.:

$$r_j(t) = (b_{jl} * n_l)(t) \qquad (3)$$

This second model can handle the effects of hardware filters within sensor channels (e.g. high-pass, notch, or antialiasing), as well as any spectral distortion or time shift that might occur within the sensor itself. The third model extends the second by allowing noise components to undergo non-linear transformations before and/or after mixing:

$$\mathbf{s}_E(t) = \mathcal{A}[\mathbf{n}(t)] \qquad (4)$$
$$\mathbf{r}(t) = \mathcal{B}[\mathbf{n}(t)]$$

where $\mathcal{A}$ and $\mathcal{B}$ are non-linear functions with memory (filters). This third model allows for non-linearities in the mixing mechanism, and/or in the sensors.

3. ALGORITHMS AND IMPLEMENTATION

Our goal is to find a function F to apply to the reference signals so that $F[\mathbf{r}(t)]$ is as close as possible to $\mathbf{s}_E(t)$ so that by subtraction we obtain the best possible measurement of $\mathbf{s}_B(t)$:

$$\widetilde{\mathbf{s}_B(t)} = \mathbf{s}(t) - F[\mathbf{r}(t)]. \qquad (5)$$

For the first model the solution is to simply project each target signal on the subspace spanned by the reference sensors. This involves finding a matrix $\mathbf{C} = [c_{kj}]$ such that the linear combination $\mathbf{Cr}(t)$ of reference signals best approximates $\mathbf{s}(t)$ in a least-squares sense. Several denoising methods proceed in this way, for example CALM [10].

For the second model (convolutive mixing), the previous solution does not work because a linear combination cannot effectively mimic or reverse the effects of convolutive mixing. We address this problem with the following algorithm (dubbed TSPCA). First, the reference channels $\mathbf{r}(t)$ are time-shifted by a series of multiples of the sampling period: $\mathbf{r}(t +$

$n), n = -N/2, \ldots, N/2$. Second, the set of time shifted references is orthogonalized by applying PCA (or other orthogonalization procedure) to obtain an orthogonal base. Components with norm below some threshold are discarded at this point to avoid numerical problems in the next step. Third, each target signal is projected on this base, and the projection removed. The result is a "clean" sensor signal. What happens, in effect, is that the combination of orthogonalization and projection forms a linear combination of time-shifted reference signals. This is the same as a N-tap FIR filter. Its coefficients are optimal, in a least squares sense, to make the reference as similar as possible to the target, and thus reverse the effects of any convolutive distortion that might differentially affect noise along each channel.

To address the third model (convolutive and nonlinear mixing), the algorithm is extended by augmenting the set of time-shifted references by non-linear transforms of the same. There are two practical approaches to choosing the non-linear transforms to be included. If the nature of distortions is known before-hand, the corresponding transforms (or their inverses if appropriate) can be included. If the nature of the distortions is not known, a selected set of nonlinearities can be provided. A linear combination of non-linear transformed reference signals can mimick a wide class of linear and non-linear distortions [11]. To the degree that this class includes or approximates the distortion actually involved, noise-related power in the target signals is reduced.

The algorithms were implemented in Matlab. Standard routines for PCA (e.g. 'svd') that operate in memory are limited in the size of the data that they can process. To lift the limit, processing is applied in several passes. The first pass calculates the covariance matrix, the second rotates and calculates the projection matrix, the third pass removes the projection from the data.

4. EVALUATION

The algorithms were tested with real and synthetic MEG data. Magnetic signals were recorded from a 160-channel, whole-head MEG system with 157 axial gradiometers sensitive to nearby sources (brain) and 3 magnetometers sensitive to distant environmental sources (KIT, Kanazawa, Japan). The system is situated within a magnetically shielded room, and sensor signals are typically filtered in hardware to reduce residual DC and very low frequency fields (highpass 1 Hz) and power line noise (notch at 60 Hz) before digitizing. Despite these precautions, when recording from the brain typically about 90% of the recorded variance is due to environmental noise. The situation is illustrated by the red plots in Fig. 1 that show the power spectrum averaged over MEG channels. In the absence of a subject (Fig. 1 (a) and (b), red) the activity reflects mainly environmental noise. The spectrum is dominated by a broad peak below 10 Hz, several narrower peaks at intermediate frequencies (10-40 Hz) and a few sharp components

at higher frequencies (≥120 Hz). Activity recorded with a subject in the machine (Fig. 1 (c), red) is dominated by the same components, and brain-related activity is therefore hard to distinguish (compare (b) and (c)).
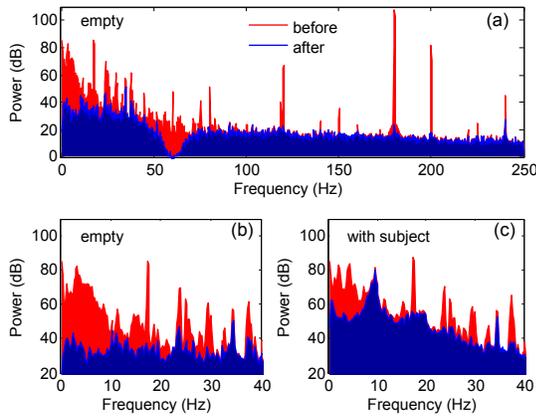
Fig. 1. (a) Power spectrum of MEG response before (red) and after (blue) denoising, averaged over all channels. There was no subject in the machine. (b) Same as (a) expanded. (c) Same as (b) with a subject in the machine.

The blue plots in Fig. 1 show the results of applying the TSPCA algorithm with $N = 200$. For data recorded with no subject (Fig. 1 (a) and (b)) the peaks at low, medium and high frequencies are suppressed. Overall, the RMS power is reduced by 98%. The result is that, in the presence of a subject, brain activity emerges more clearly (Fig. 1 (c)). Assuming that brain and noise activity are decorrelated, the SNR can be estimated approximately by comparing activity with and without a subject. After denoising, the SNR is better than 15 dB over the 0-20 Hz range important for brain activity, with a peak of more than 40 dB at 9 Hz (not shown).
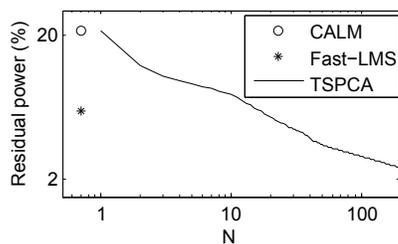
Fig. 2. Percentage of residual power after TSPCA as a function of the number of taps N. Also shown are figures for two other algorithms: CALM [10] and Fast-LMS [12]
.

The effectiveness of TSPCA stems from its ability to compensate for a convolutional mismatch between target and reference channels, as illustrated by Fig. 2. For $N = 1$ the algorithm degenerates to simple scalar denoising, and is in fact equivalent to the CALM algorithm (circle). With $N = 200$, TSPCA offers an order of magnitude less noise than CALM. It surpasses also the Fast-LMS algorithm of [12].

An important concern of scientists and clinicians is to make sure that processing (such as denoising) does not distort components of interest. MSPCA involves filtering but, as Eq. 5 shows, it is applied only to the reference and not the target signals. It is true that targets are stripped of any components that happen to be colinear with the subspace spanned by the (time-shifted) reference channels. However, if brain activity is not correlated with environmental noise, these components should be small. This conclusion was tested by simulating the denoising situation with synthetic data. For convenience we used gaussian noise (uncorrelated across channels) as a "target" signal. For "noise" we used activity recorded from an MEG in the absence of a subject, stripped of the residual after denoising (2% of its variance). This "noise" is essentially similar to real environmental noise, but with the convenient property that denoising removes it perfectly, so that target distortion is easier to assess. As evident in Fig. 3 (compare blue and green traces), the only deleterious effect of denoising is a very slight reduction in amplitude. Simulations show that the amplitude loss is proportional to N (not shown), suggesting that the effect is the result of increased overfitting as N increases. TSPCA does *not* significantly distort the target signals.
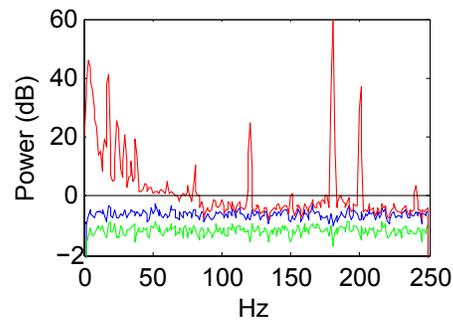
Fig. 3. Effect of denoising on target illustrated with synthetic data. Blue: power spectrum of synthetic "target" before adding noise. Red: power spectrum of "noise". Green: power spectrum of target after denoising ($N = 200$). Note that negative ordinates are expanded (the 0 dB level is arbitrary).
.

TSPCA addresses signal model 2 (convolutive mixing). To address model 3 (convolutive and non-linear), the same algorithm can be extended by including a set of nonlinear transforms of the delayed reference channels. A wide class of nonlinearities can be approximated as weighted sums of a much smaller class of nonlinear functions such as polynomials [11]. Including just a few appropriately-chosen nonlinear transforms is expected to improve the fit of the noise component of target channels to their projection on the reference-derived subspace. For real MEG data, it turns out that this does *not* significantly reduce noise variance, presumably because the mixing and transduction process is highly linear.

However other application scenarios might involve some degree of nonlinearity. To test the effectiveness of the algorithm in the non-linear case, we synthesized artificial data for which the noise components affecting the target $(\mathbf{s}_E(t))$ and those observable from the reference $(\mathbf{r}(t))$ were related by a non-linear function (square root). In this case, standard (linear) TSPCA is ineffective. However, appending a set of non-linear transformed channels (powers) improves noise suppression; and indeed applying the *same* nonlinearity (square root) allows noise suppression as effective as in the linear case (not shown).

5. DISCUSSION AND CONCLUSION

The TSPCA algorithm is remarkably effective in removing environmental magnetic noise from MEG recordings, and it seems that the same method should work well for other physiological recording techniques, such as EEG, local field potentials, etc. A necessary condition for applying TSPCA is the availability of reference channels sensitive only to noise. Other algorithms exist that remove noise based on such channels. The original feature of TSPCA is that it can handle a convolutive mismatch between target and reference channels, as well as various forms of nonlinearity. This last feature is tested only superficially here: future work will attempt to characterize these nonlinear-denoising capabilities in greater depth. The ability to effectively suppress high-levels of environmental noise is crucial to the deployment of MEG systems in health applications, as high-quality shielding is expensive and bulky. For a given environment it can lead to better-quality data, and for scientific investigations it reduces the need for long experiments, involving multiple presentation of the same stimulus. Noise immunity is also an important step towards practical brain-machine interfaces.

Acknowledgements

6. REFERENCES

[1] A. de Cheveigné and J.Z. Simon, "Magnetoencephalography denoising based on time-shift PCA," *J. Neurophysiology*, 2007 (submitted).

[2] M. Hamalainen, R. Hari, P. J Ilmoniemi, J. K. Uutila, and O. V. Lounasmaa, "Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain," *Rev. Mod. Phys.*, vol. 65, pp. 413–497, 1993.

[3] J. Vrba, "Multichannel SQUID biomagnetic systems," *NATO ASI Series: E Applied Sciences*, vol. 365, pp. 61–138, 2000.

[4] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," *IEEE Signal Processing Magazine*, vol. 18, pp. 14–30, 2001.

[5] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda, "Recipes for the linear analysis of EEG," *NeuroImage*, vol. 28, pp. 326–341, 2005.

[6] K. M. Spencer, J. Dien, and E. Donchin, "Spatiotemporal analysis of the late ERP responses to deviant stimuli," *Psychophysiology*, vol. 38, pp. 343–358, 2001.

[7] E. M. Bernat, W. J. Williams, and W. J. Gehring, "Decomposing ERP time-frequency energy using PCA," *Clinical Neurophysiology*, vol. 116, pp. 1314–1334, 2005.

[8] E. Ahissar, S. Nagarajan, M. Ahissar, A. Protopapas, H. Mahncke, and M.M. Merzenich, "Speech comprehension is correlated with temporal response patterns recorded from auditory cortex," *Proceedings of the National Academy of Science (U.S.A.)*, vol. 98, pp. 13367–13372, 2001.

[9] J. Kayser and C. E. Tenke, "Optimizing PCA methodology for ERP component identification and measurement: theoretical rationale and empirical evaluation," *Clinical Neurophysiology*, vol. 114, pp. 2307–2325, 2003.

[10] Y. Adachi, M. Shimogawara, M. Higuchi, Y. Haruta, and M. Ochiai, "Reduction of non-periodic environmental magnetic noise in MEG measurement by continuously adjusted least squares method," *EEE Trans on Applied Superconductivity*, vol. 11, pp. 669–672, 2001.

[11] S. Haykin, "Neural networks, a comprehensive foundation," Upper Saddle River,New Jersey : Prentice Hall, 2001.

[12] N. Ahmar and J.Z. Simon, "MEG adaptive noise suppression using fast LMS," in *2nd International IEEE EMBS Conference on Neural Engineering*, 2005, pp. 29–32.