BLIND SOURCE SEPARATION APPROACHES TO REMOVE IMAGING ARTEFACTS IN EEG SIGNALS RECORDED SIMULTANEOUSLY WITH fMRI

Bertrand Rivet\textsuperscript{1,2}, Guillaume Flandin\textsuperscript{3}, Antoine Souloumiac\textsuperscript{2}, Jean-Baptiste Poline\textsuperscript{3}

\textsuperscript{1} GIPSA-Lab, CNRS-UMR 5216, Grenoble Institute of Technology
\textsuperscript{2} CEA, LIST, Laboratoire Processus Stochastiques et Spectres, CEA, Saclay, GIF sur Yvette, F-91191, France
\textsuperscript{3} CEA, Neurospin, I2BM, Saclay, CEA, Saclay, GIF sur Yvette, F-91191, France

ABSTRACT

Using jointly functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) is a growing field in human brain mapping. However, EEG signals are contaminated during acquisition by imaging artefacts which are stronger by several orders of magnitude than the brain activity. In this paper, we propose three methods to remove the imaging artefacts based on the temporal and/or the spatial structures of these specific artefacts. Moreover, we propose a new objective criterion to measure the performance of the proposed algorithm on real data. Finally, we show the efficiency of the proposed methods applied to a real dataset.

1. INTRODUCTION

The combination of electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) has recently been investigated in human brain imaging [1, 2, 3]. The fMRI modality provides signals related to the hemodynamic neuronal activity with a very high spatial resolution (around $2 \times 2 \times 2$ mm$^3$) and with a low temporal resolution (around 3s). A contrario, the EEG modality provides signals related to the electrophysiological activity with a very high temporal resolution (around 1kHz) and with a low spatial resolution (from 32 to 512 scalp sensors). As a consequence, some studies investigated the possibility of using the strengths of these two techniques by combining their complementarities [2].

However, the EEG signals recorded during MRI acquisition contain two main types of artefacts due to the magnetic field used by the MRI scanner: the ballistocardiogram (BCG) and imaging artefacts. BCG artefact is related to cardiac rhythm and is mainly due to the heart-related blood and electrodes movements in the magnetic field. Imaging artefact is induced by the gradient magnetic fields used for spatial encoding in MRI. Different methods were proposed to attenuate these artefacts, see [1, 4, 3] for instance. They exploit separately the temporal structure of imaging artifact and/or the spatial structure of imaging. In this paper, we address the same problem of removing imaging artifact. The proposed methods also exploit the temporal and spatial structures of the imaging artifact but in different ways.

This paper is organized as follows. Section 2 describes the temporal and spatial structures of imaging artifact as well as the proposed methods to remove them. Section 3 presents the results that have been achieved whereas Section 4 concludes the paper with comments and perspectives.

2. IMAGING ARTEFACT REMOVAL

In this section, the temporal and spatial structures of imaging artefact are stressed and we explain how we propose to exploit them to remove imaging artefact in EEG signals.

The two main properties rest upon the fact that the imaging artifact reflects the switching of gradient magnetic field used to record MRI where a volume is composed of several slices, each of them representing an fMRI image. Firstly, since during a classical fMRI recording each volume is composed of identical slices, the associated gradient magnetic field is the same for each volume. Secondly, since all the EEG sensors are immersed in the same magnetic field, they record the same physical phenomenon in different ways.

As a consequence, on the first hand the effect of recording different volumes must have the same influence in the recorded EEG during the experiment, and on the other hand the imaging artefact must occupy a small spatial subspace of space spanned by the recorded EEG.

2.1 Temporal model of imaging artefact

Let $x_i(t)$ denote the EEG signal recorded by the $i$-th sensor at continuous time $t$. As explained above, the influence of artefact gradient may be modelled by

$$x_i(t) = \sum_j g_i(t - \tau_j) + e_i(t),$$

where $\tau_j$ is the $j$-th volume timing event, $g_i(t)$ is a function which expresses the imaging artefacts of one volume on the $i$-th sensor, and $e_i(t)$ is the term of ongoing brain activity. A classical approach to estimate $e_i(t)$ is first to estimate $g_i(t)$ and then to remove it from $x_i(t)$:

$$\hat{e}_i(t) = x_i(t) - \sum_j \hat{g}_i(t) * \delta(t - \tau_j),$$

where $*$ is the convolution product and $\delta(t)$ is the delta Dirac function. Under the assumption that EEG activity is uncorrelated if the time delay is larger than $\min(\tau_{n+1} - \tau_n)$, one can estimate $\hat{g}_i(t)$ for each sensor by

$$\hat{g}_i(t) = \frac{1}{N} \sum_{k=0}^{N-1} x_i(t + \tau_k).$$

However, since we are dealing with discrete time signals, (3) becomes

$$\hat{g}_i[n] \triangleq \hat{g}_i(nT_s) = \frac{1}{N} \sum_{k=0}^{N-1} x_i\left( n + \frac{\tau_k}{T_s} \right) T_s$$

where $T_s$ is the sampling period. The main difficulty comes from the asynchronously clocks of EEG and fMRI data: as a consequence $\tau_k/T_s$ is not necessary an integer, thus substituting $\tau_k/T_s$ by $n_k$, the closest integers to $\tau_k/T_s$, leads to an
awkward estimation of the imaging artefact. A common solution [1, 4, 3] is to interpolate the EEG data by oversampling to obtain a better estimation of \( \tau_k \). However, this solution suffers from two main drawbacks: over-sampled data require more memory and a high resampling rate is needed to obtain a good alignment. To overcome this, we propose to estimate \( \tau_k \) and to time shift \( x(t) \) without any oversampling thanks to the following property

\[
\text{TF}(x(t - \tau_k)) = X(f) \exp(-i \pi \tau_k f),
\]

(5)

where TF(\cdot) is the Fourier transform operator, \( \tau^2 = -1 \) and \( X(f) \) is the Fourier transform of \( x(t) \).

Let \( x^{(k)}_i = [x_i[n_k], \ldots, x_i[n_k + N - 1]]^T \) be EEG signal during the acquisition of \( k \)-th volume and \( X^{(k)}_i = [X_i[0], \ldots, X_i[N - 1]]^T \) its Fourier transform. Thus the inter-spectrum of \( X^{(0)}_i[v] \) and \( X^{(k)}_i[v] \) for all \( k \) is defined by

\[
X^{(0)}_i[v] \left( X^{(k)}_i[v] \right)^* = \left( G^{(0)}_i[v] + E^{(0)}_i[v] \right) \times \left( G^{(k)}_i[v] + E^{(k)}_i[v] \right)^*,
\]

where \(^*\) is the complex conjugate. Since the imaging artefact is much stronger than the ongoing brain activity and thanks to (5), and choosing arbitrary \( \tau_0 = 0 \), the inter-spectrum can be expressed as

\[
X^{(0)}_i[v] \left( X^{(k)}_i[v] \right)^* \simeq \left| G^{(0)}_i[v] \right|^2 \exp \left( i 2 \pi \frac{v}{N} \tau'_k \right),
\]

(6)

where \( \tau'_k = \frac{\tau_k}{2} - n_k \). One can see that the phase of the inter-spectrum depends linearly on \( \tau'_k \). Note that \( n_k \) is known thanks to the triggers received from the MRI machine which indicates the start of each volume. \( \tau'_k \) is thus estimated by a linear regression on the phase for frequency bins in the pass-band of \( g_i(t) \) defined by frequencies whose modulus amplitude \( \left| G^{(0)}_i[v] \right|^2 \) represents a definite part of the power of \( g_i(t) \) approximated by the power of \( x_i(t) \) (typically more than 10%).

Finally, the imaging artefact is estimated by

\[
\hat{g}_i[n] = \frac{1}{K} \sum_{k=0}^{K-1} X^{(k)}_i[n],
\]

(7)

where \( X^{(k)}_i[n] \) is the re-aligned observations obtained by the inverse Fourier transform of \( X^{(k)}_i[v] \exp(2\pi \nu \tau'_k / N) \), and the ongoing brain activity \( x_i[n] \) is estimated by

\[
\hat{x}_i[n] = x_i[n] - \sum_j \hat{g}_j^{(j)}[n],
\]

(8)

where \( \hat{g}_j^{(j)}[n] \) is the inverse Fourier transform of \( G^{(j)}_j[v] \exp(-2\pi \nu \tau'_j / N) \). This algorithm to cancel the influence of imaging artefact on EEG signal is called Frequency Averaged Artefact Subtraction (F-AAS) and is resumed in Algorithm 1.

### 2.2 Spatial model of imaging artefact

The fact that all the sensors record the same physical phenomenon in different ways can be modeled by

\[
x[n] = A g[n] + e[n],
\]

(9)

where \( x[n] = [x_1[n], \ldots, x_N[n]]^T \) is the column vector of the recorded signals, \( g[n] = [g_1[n], \ldots, g_N[n]]^T \) is the column vector expressing the imaging artefact and \( e[n] = [e_1[n], \ldots, e_N[n]]^T \) is the column vector of the ongoing brain activity. \( A \in \mathbb{R}^{N_t \times N_g} \) is the mixing matrix. Model (9) is thus a linearly instantaneous mixture which can be inverted by independent component analysis (ICA) [5]). It aims at finding a separation matrix \( B \) such that \( y[n] = B x[n] \) is a vector with mutually independent components. Most of used ICA algorithms for EEG signal processing are based on non-Gaussianity [6, 7] or based on time coherence [8]. To estimate separation matrix \( B \), we propose to exploit the non-stationarity of the imaging artefact: as can be seen on Fig. 3(a), the imaging artefact on EEG signals is only present during the recording of volumes. If the sources are assumed to be mutually independent (or at least uncorrelated), covariance matrices \( C_{xy}(n) \) of signal \( y[n] \) at several time indexes \( n \) must be diagonal. A basic criterion for blind source separation (BSS) [9] is to compute \( C_{xy}(n) \) from the observations \( x[n] \) and then to adjust matrix \( B \) such that \( C_{xy}(n) \) is as diagonal as possible. Since this condition must be verified for any time index \( n \), this can be done by a joint-diagonalisation method, and in the following we use the algorithm of [10].

Algorithm 1 F-AAS algorithm.

1: for each sensor \( i \) do
2: Compute Fourier transform of \( x_i^{(k)}[n] \Rightarrow \hat{X}_i^{(k)}[v] \)
3: for each volume \( k \) do
4: Compute inter-spectrum (6): \( \chi_i^{(k)}[v] \left( \hat{X}_i^{(k)}[v] \right)^* \)
5: Estimate time delay \( \tau'_k \) by linear regression on inter-spectrum phase
6: Temporal alignment of \( x^{(k)}[n] \Rightarrow \tilde{x}^{(k)}[n] \)
7: end for
8: Estimate imaging artefact \( \tilde{g}_i[n] \) by (7)
9: Estimate brain activity \( \hat{e}^{(k)}[n] \) by (8)
10: end for

where \( x[n] = [x_1[n], \ldots, x_N[n]]^T \) is the column vector of the recorded signals, \( g[n] = [g_1[n], \ldots, g_N[n]]^T \) is the column vector expressing the imaging artefact and \( e[n] = [e_1[n], \ldots, e_N[n]]^T \) is the column vector of the ongoing brain activity. \( A \in \mathbb{R}^{N_t \times N_g} \) is the mixing matrix. Model (9) is thus a linearly instantaneous mixture which can be inverted by independent component analysis (ICA) [5]). It aims at finding a separation matrix \( B \) such that \( y[n] = B x[n] \) is a vector with mutually independent components. Most of used ICA algorithms for EEG signal processing are based on non-Gaussianity [6, 7] or based on time coherence [8]. To estimate separation matrix \( B \), we propose to exploit the non-stationarity of the imaging artefact: as can be seen on Fig. 3(a), the imaging artefact on EEG signals is only present during the recording of volumes. If the sources are assumed to be mutually independent (or at least uncorrelated), covariance matrices \( C_{xy}(n) \) of signal \( y[n] \) at several time indexes \( n \) must be diagonal. A basic criterion for blind source separation (BSS) [9] is to compute \( C_{xy}(n) \) from the observations \( x[n] \) and then to adjust matrix \( B \) such that \( C_{xy}(n) \) is as diagonal as possible. Since this condition must be verified for any time index \( n \), this can be done by a joint-diagonalisation method, and in the following we use the algorithm of [10].

Among the estimated sources \( y_i[n] \), we proposed to select those which contain imaging artefact: let \( I \subseteq \{1, \ldots, N_g\} \) denote this set of \( N_g \) indexes. Then, the F-AAS algorithm is applied on each source \( y_i[n] \) if \( i \in I \), and each source \( y_i[n] \) such that \( i \notin I \) are kept unaltered. Finally, the brain activity is estimated by

\[
\hat{e}[n] = B^{-1} y'[n],
\]

(10)

where \( y'[n] \) is the vector composed of the unselected estimated sources \( y_i(t) \), with \( i \notin I \) plus the estimated sources \( y_i(t) \), with \( i \notin I \) denoised by F-AAS algorithm.

We call this procedure Spatial Averaged Artefact Subtraction (S-AAS), and it is summarized in Algorithm 2. Note that this second approach seems more conservative than F-AAS applied directly on all sensors \( x[n] \). Indeed, F-AAS algorithm may result in the modification of brain activity EEG data. However to express EEG data as \( x[n] \) in the sensor space (9) or to express them as \( y[n] \) in the source space is equivalent since estimated matrix \( B \) is invertible. Moreover the imaging artefact results from the magnetic field thus its dimension must be reduced compared to the numbers of EEG sensors. As a result, ICA may concentrate the influence of this magnetic field in a limited number of components \( N_g < N_e \). S-AAS algorithm finally keeps unaltered \( N_e - N_g \).
signals, minimizing the eventual awkward influences of F-AAS.

Algorithm 2 S-AAS algorithm.
1: Compute a set of covariance matrices \( \{C_{xx}(n)\}_n \).
2: Estimate matrix \( B \) by joint-diagonalization of \( \{C_{xx}(n)\}_n \).
3: Compute estimated sources: \( y[n] = B \hat{x}[n] \).
4: Select sources contaminated by imaging artefact \( \Rightarrow \mathcal{F}_g \).
5: for each source \( i \) do
6: if \( i \in \mathcal{F}_g \) then
7: Apply F-AAS algorithm on \( y_i[n] \Rightarrow y_i'[n] \) obtained by (8).
8: else
9: Keep unaltered \( y_i[n] \Rightarrow y_i'[n] = y_i[n] \).
10: end if
11: end for
12: Estimate brain activity \( \hat{\theta}[n] \) by (10).

2.3 Spatio-temporal model of imaging artefact

As explained in Subsections 2.1 and 2.2, the imaging artefact has a temporal structure and a spatial structure. This spatio-temporal structure can then be modelled as a convolutive mixture
\[
x(t) = A(t) \ast g(t) + e(t),
\]
where \( A(t) \) is the mixing filter matrix whose \((k,l)\)-th entry is expressed as \( \sum_{i}A_{i,k}^l \delta(t - \tau_i) \): \( A(t) = \sum_{i}A_{i}^l \delta(t - \tau_i) \). One can then estimate \( \hat{g}(t) \) thanks to a separating filter matrix \( B(t) \) by
\[
\hat{g}(t) = B(t) \ast x(t),
\]
where \( B(t) \) can be expressed as
\[
B(t) = \sum_j B_j \delta(t + \tau_j).
\]

First, to overcome the problem of asynchronous clocks between EEG and MRI data, the same estimation of \( \tau_j \) than proposed in Subsection 2.1 is used. Second, to enforce the impulse response of \( B(t) \) to have the special structure (13), the observations \( \hat{x}^{(k)}[n] \) are first computed and time shifted to obtain \( \hat{x}^{(k)}[n] \). Then block Toeplitz matrix \( Z \in \mathbb{R}^{(N+J) \times (NK)} \) is computed such that
\[
Z = \begin{bmatrix}
\hat{x}_1^{(0)T} & \hat{x}_1^{(1)T} & \cdots & \hat{x}_1^{(K-1)T} \\
\vdots & \ddots & \vdots & \vdots \\
\hat{x}_N^{(0)T} & \hat{x}_N^{(1)T} & \cdots & \hat{x}_N^{(K-1)T} \\
\vdots & \ddots & \vdots & \vdots \\
\hat{x}_1^{(J-1)T} & \hat{x}_1^{(J)T} & \cdots & \hat{x}_1^{(J+K-2)T} \\
\vdots & \ddots & \vdots & \vdots \\
\hat{x}_N^{(J-1)T} & \hat{x}_N^{(J)T} & \cdots & \hat{x}_N^{(J+K-2)T}
\end{bmatrix},
\]
where \( \hat{x}_i^{(k)} = [\hat{x}_i^{(k)}[0], \ldots, \hat{x}_i^{(k)}[N - 1]]^T \). Finally, separating filters matrix \( B(t) \) is estimated thanks to the non-stationary blind source separation method presented in Subsection 2.2.

Algorithm 3 ST-AAS algorithm.
1: for each volume \( k \) do
2: Temporal alignment of \( x^{(k)}[n] \Rightarrow \hat{x}^{(k)}[n] \).
3: end for
4: Compute matrix \( Z \) (14).
5: PCA of \( Z \Rightarrow W \) and \( Z' \).
6: Compute a set of covariance matrices \( \{C_{xZ'}(n)\}_n \).
7: Estimate \( R \) by joint-diagonalisation.
8: Select sources contaminated by imaging artefact \( \Rightarrow \mathcal{F}_g \).
9: Estimate brain activity \( \hat{\theta}[n] \) by (16).

3. RESULTS

In this section, the data acquisition process and the results obtained by the proposed methods are presented.

3.1 Data acquisition

EEG was acquired using the MRI-compatible BrainAmp MR (BrainProducts, Munich, Germany) EEG amplifier and the BrainCap electrode cap (EasyCap, Herrsching-Breitbrunn, Germany) with sintered Ag/AgCl non-magnetic ring electrodes providing 32 EEG channels. They were positioned according to the classic 10-20 system. Raw EEG was sampled at 5kHz using the BrainVision Recorder software (BrainProducts) with a signal range +/-16mV (16-bit sampling). A 7-minute session was recorded inside the MRI environment during image acquisition while the subject was presented a visual stimulation (flashing rings with a perceived frequency of 5Hz). EEG data were hardware filtered using a low-pass filter (fc=250 Hz), a high-pass filter (fc=0.016Hz) and a notch-filter around 50Hz. Functional MRI images were acquired on a 3T Trio scanner (Siemens, Erlangen, Germany) with a standard head coil using an Echo Planar Imaging (EPI) sequence covering almost the entire brain (TR=45ms, 1.5s gap, TE=45, 64 axial slices, voxel size 3.75 x 3.75 x 5 mm3, 2 mm gap). This sparse acquisition allows us to have a 1.5 second window gradient artefact free followed by a 1.5 second of artefacted signal for each fMRI volume. Importantly, there was no synchronisation between the MR sequence and the EEG amplifier clocks.

Before applying the proposed methods, the raw EEG signals were first 1Hz high-pass filtered by a 2nd order forward-forward Butterworth filter. After the estimation of time
One can see that the influence of the imaging gradient is different for the low-pass filtered observations \( Z_t \) (GEV) between MRI acquisitions. And let finally evaluate the estimations. Let \( n > 1 \) \([11]\). We thus propose a new criterion to evaluate the arterial artefact have a spatial structure which can be enhanced by a spatial filtering. The F-AAS algorithm applied on these specific components efficiently removes the imaging artefact (Fig. 2(a)) which is confirmed by the final result shown in Fig. 3(c) and 3(g). In this experiment and in the following, the sources which contain imaging artefacts are selected manually. Actually this selection could be done automatically with a short term power criterion for instance.

In a second experiment, the S-AAS algorithm (Subsection 2.2) was applied on the data. One can see on Fig. 2(a) that the blind source separation concentrates the imaging artefact in a limited number of components (nine indicated by arrows). This fact confirms once again that the imaging artefact have a spatial structure which can be enhanced by a spatial filtering. The F-AAS algorithm applied on these specific components efficiently removes the imaging artefact (Fig. 2(a)) which is confirmed by the final result shown in Fig. 3(c) and 3(g). In this experiment and in the following, the sources which contain imaging artefacts are selected manually. Actually this selection could be done automatically with a short term power criterion for instance.

In the last experiment, the ST-AAS algorithm (Subsection 2.3) was applied on the data. Matrix \( Z \) was computed with \( J = 38 \) and \( K = 5 \) leading to a 1216 \( \times 65000 \) matrix \( Z \) since \( N = 13000 \). During the PCA stage, matrix \( Z \) is reduced to a 15 \( \times 65000 \) matrix \( Z' \), leading thus to 15 estimated sources \( \hat{y}_i[n] \) plotted in Fig. 2(b). One can see that most of the imaging artefact is concentrated in four sources (indicated by arrows) which were thus removed from the observations \( x[n] \). The final result of the ST-AAS is shown in Fig. 3(d) and 3(h). Quite surprisingly, this algorithm seems to have less performance than F-AAS and S-AAS. This might be explained by the fact that this algorithm is a bit more difficult to use since it needs to fix more parameters \((J, K, J')\) the number of principal components kept).

Finally, the three proposed algorithms seems to have quite good performance. On the first hand, F-AAS and S-AAS have a similar computational cost but, as already mentioned (Subsection 2.2), S-AAS seems more conservative and thus seems better (Fig. 3(f) and 3(g)). The two sets \( P_1 \) and \( P_2 \) are closer using S-AAS algorithm than using F-AAS. On the other hand, the ST-AAS has a higher computational cost and has little bit less good performance (Fig. 3(h)). Indeed, one can see that three largest GEV are quite different.
methods have to be validated on other databases. Moreover, even if the evaluation on real data is an open problem, the proposed objective criterion seems to be quite efficient on real data. Finally, to be complete, the proposed algorithms, which exploit separately the spatial and temporal structures, seem to show that exploiting jointly spatial and temporal structures should be relevant. We are currently investigating new implementation to take into account the spatio-temporal structure of imaging artefact to improve the ST-AAS algorithm.

Moreover, even if the evaluation on real data is an open problem, the proposed objective criterion seems to be quite efficient to measure the performance of imaging artefact removal on real data. Finally, to be complete, the proposed methods have to be validated on other databases.

REFERENCES


