

RESEARCH

Open Access

Separation of phase-locked sources in pseudo-real MEG data

Miguel Almeida^{1*}, José Bioucas-Dias¹ and Ricardo Vigário²

Abstract

This article addresses the blind separation of linear mixtures of synchronous signals (i.e., signals with locked phases), which is a relevant problem, e.g., in the analysis of electrophysiological signals of the brain such as the electroencephalogram and the magnetoencephalogram (MEG). Popular separation techniques such as independent component analysis are not adequate for phase-locked signals, because such signals have strong mutual dependency. Aiming at unmixing this class of signals, we have recently introduced the independent phase analysis (IPA) algorithm, which can be used to separate synchronous sources. Here, we apply IPA to pseudo-real MEG data. The results show that this algorithm is able to separate phase-locked MEG sources in situations where the phase jitter (i.e., the deviation from the perfectly synchronized case) is moderate. This represents a significant step towards performing phase-based source separation on real data.

1 Introduction

In recent years, the interest of the scientific community in synchrony has risen. This interest is both in its physical manifestations and in the development of a theory unifying and describing those manifestations in various systems such as laser beams, astrophysical objects, and brain neurons [1].

It is believed that synchrony plays a relevant role in the way different parts of the human brain interact. For example, when humans engage in a motor task, several brain regions oscillate coherently [2,3]. Also, several pathologies such as autism, Alzheimer, and Parkinson are associated with a disruption in the synchronization profile of the brain, whereas epilepsy is associated with an anomalous increase in synchrony (see [4] for a review).

To perform inference on the synchrony of networks present in the brain or in other real-world systems, one must have access to the phase dynamics of the individual oscillators (which we will call “sources”). Unfortunately, in brain electrophysiological signals such as encephalograms (EEG) and magnetoencephalograms (MEG), and in other real-world situations, individual oscillator signals are not directly measurable, and one has only access to a superposition of the sources^a. In fact, EEG and MEG signals

measured in one sensor contain components coming from several brain regions [5]. In this case, spurious synchrony may occur, as we will illustrate later.

The problem of undoing this superposition is called blind source separation (BSS). Typically, one assumes that the mixing is linear and instantaneous, which is a valid approximation in brain signals [6]. One must also make some assumptions on the sources, such as in independent component analysis (ICA) where the assumption is mutual statistical independence of the sources [7]. ICA has seen multiple applications in EEG and MEG processing (for recent applications see, e.g., [8,9]). Different BSS approaches use criteria other than statistical independence, such as non-negativity of sources [10,11] or time-dependent frequency spectrum criteria [12,13]. In our case, independence of the sources is not a valid assumption, because phase-locked sources are highly mutually dependent. Also, phase-locking is not equivalent to frequency coherence: in fact, two signals may have a severe overlap between their frequency spectra but still exhibit low or no phase synchrony at all [14]. In this article, we address the problem of how to separate such phase-locked sources using a phase-specific criterion.

Recently, we have presented a two-stage algorithm called independent phase analysis (IPA) which performed very well in noiseless simulated data [15] and with moderate levels of added Gaussian white noise [14]. The

*Correspondence: malmeida@lx.it.pt

¹Instituto de Telecomunicacoes, Lisbon, Portugal

Full list of author information is available at the end of the article

separation algorithm we then proposed uses temporal decorrelation separation [16] as a first step, followed by the maximization of an objective function involving the phases of the estimated sources. In [14], we presented a “proof-of-concept” of IPA, laying down the theoretical foundations of the algorithm and applying it to a toy dataset of manually generated data. However, in that article we were not concerned with the application of IPA to real-world data. In this article, we study the applicability of IPA to pseudo-real MEG data. These data are not yet meant to allow inference about the human brain; however, they are generated in such a way that both the sources and the mixing process mimic what actually happens in the human brain. The advantage of using such pseudo-real data is that the true solution is known, thus allowing a quantitative assessment of the performance of the algorithm. We also study the robustness of IPA to the case where the sources are not perfectly phase-locked. It should however be reinforced that the algorithm presented here makes no assumptions that are specific of brain signals, and should work in any situation where phase-locked sources are mixed approximately linearly and noise levels are low.

This article is organized as follows. In Section 2, we introduce the Hilbert transform. We also introduce there the phase locking factor (PLF), a measurement of synchrony which is central to the algorithm; finally, we show that synchrony is disrupted when the sources undergo a linear mixing. Section 3 describes the IPA algorithm in detail, including illustrations using a toy dataset. In Section 4, we explain how the pseudo-real MEG data are generated and show the results obtained by IPA on those data. These results are discussed in Section 5 and conclusions are drawn in Section 6.

2 Background

2.1 Hilbert transform: phase of a real-valued signal

Usually, the signals under study are real-valued discrete signals. To obtain the phase of a real signal, one can use a complex Morlet (or Gabor) wavelet transform, which can be seen as a bank of bandpass filters [17]. Alternatively, one can use the Hilbert transform, which should be applied to a locally narrowband signal or be preceded by appropriate filtering [18] for the meaning of the phase extracted by the Hilbert transform to be clear. The two transforms have been shown to be equivalent for the study of brain signals [19], but they may differ for other kinds of signals. In this article, we chose to use the Hilbert transform. To ensure that this transform yields meaningful results, we will precede its use by band-pass filtering the pseudo-real MEG sources used in this article (see Section 4.1). Note that this is a very common preprocessing step in the analysis of real MEG signals (cf., [20-22]).

The discrete Hilbert transform $x_h(t)$ of a band-limited discrete-time signal $x(t)$, $t \in \mathbb{Z}$, is given by a convolution [18]

$$x_h(t) \equiv x(t)*h(t), \text{ where } h(t) \equiv \begin{cases} 0, & \text{for even } t \\ \frac{2}{\pi t}, & \text{for odd } t. \end{cases}$$

Note that the Hilbert transform is a linear operator. The Hilbert filter $h(t)$ is not causal and has infinite duration, which makes direct implementation of the above formula impossible. In practice, the Hilbert transform is usually computed in the frequency domain, where the above convolution becomes a product of the discrete Fourier transforms of $x(t)$ and $h(t)$. A more thorough mathematical explanation of this transform is given in [18,23]. We used the Hilbert transform as implemented by MATLAB.

The analytic signal of $x(t)$, denoted by $\tilde{x}(t)$, is given by $\tilde{x}(t) \equiv x(t) + i x_h(t)$, where $i = \sqrt{-1}$ is the imaginary unit. The phase of $x(t)$ is defined as the angle of its analytic signal. In the remainder of the article, we drop the tilde notation; it should be clear from the context whether the signals under consideration are the real signals or the corresponding analytic signals.

2.2 Phase-locked sources

Throughout this article, we assume that the sought sources, in number of N and denoted by s_j , $j = 1, \dots, N$, are phase-locked. In other words, s_j , $j = 1, \dots, N$ are complex valued signals with nonnegative amplitudes and equal phase up to a constant plus small perturbations. Formally,

$$s_j(t) = a_j(t)e^{i(\alpha_j + \phi(t) + \delta_j(t))}, \quad (1)$$

where $a_j(t)$ are the amplitudes of the sources, which are by definition non-negative and real-valued. α_j is the constant dephasing (or phase lag) between the sources (it does not depend on the time t), $\phi(t)$ represents an oscillation common to all the sources (it does not depend on the source j), and $\delta_j(t)$ is the phase jitter, which represents the deviation of the j th source from its nominal phase $\alpha_j + \phi(t)$. Throughout this article, we will assume that the phase jitter is Gaussian with zero mean and a standard deviation σ .

One situation where signals follow the model in (1) is the one described by the (time-dependent) Kuramoto model, under some circumstances. This simple model has extensively been used in the context of, e.g., modeling neuronal excitation and inhibition interactions, as well as large-scale experimental neuroscience data [20,24]. Under this model, the interactions between oscillators are weak relative to the stability of their limit cycles, and thus affect the

oscillators' phases only, not their amplitudes. The phase of oscillator j is governed by [1,25,26]

$$\dot{\phi}_j(t) = \omega_j(t) + \frac{1}{N} \sum_{k=1}^N \kappa_{jk} \sin [\phi_k(t) - \phi_j(t)], \quad (2)$$

where $\phi_j(t)$ is the phase of oscillator j (it is unrelated to $\phi(t)$ in Equation (1)), $\omega_j(t)$ is its natural frequency, and κ_{jk} measures the strength of the interaction between oscillators j and k . If the κ_{jk} coefficients are large enough and $\omega_j(t) = \omega_k(t)$ for all j, k , then the solutions of the Kuramoto model are of the form (1) with small $\delta_j(t)$.

2.3 PLF

Given two oscillators with phases $\phi_j(t)$ and $\phi_k(t)$ for $t = 1, \dots, T$, the real-valued^b PLF, or phase locking value, between those two oscillators is defined as

$$\varrho_{jk} \equiv \left| \frac{1}{T} \sum_{t=1}^T e^{i[\phi_j(t) - \phi_k(t)]} \right| = \left| \left\langle e^{i(\phi_j - \phi_k)} \right\rangle \right|, \quad (3)$$

where $\langle \cdot \rangle$ is the time average operator. The PLF satisfies $0 \leq \varrho_{jk} \leq 1$. The value $\varrho_{jk} = 1$ corresponds to two oscillators that are fully synchronized (i.e., their phase lag is constant). In terms of Equation (1), a PLF of 1 is obtained only if the phase jitter $\delta_j(t)$ is zero. The value $\varrho_{jk} = 0$ is attained, for example, if the phase difference $\phi_j(t) - \phi_k(t)$ modulo 2π is uniformly distributed in $[-\pi, \pi]$. Values between 0 and 1 represent partial synchrony; in general, higher values of the standard deviation of the phase jitter $\delta_j(t)$ yield lower PLF values.

Note that a PLF of 1 is obtained if and only if $\phi_j(t) - \phi_k(t)$ is constant^c. Thus, studying the separation of sources with constant phase lags can equivalently become the study of separation of sources with pairwise PLFs of 1.

Throughout this article, phase synchrony is measured using the PLF; two signals are perfectly synchronous if and only if they a PLF of 1. Other approaches exist, e.g., for chaotic systems or specific types of oscillators [27]. Studying separation algorithms based on such other definitions is outside of the scope of this article. The definition used here has the advantages of being tractable from an algorithmic point of view, and of being applicable to any situation where $\phi_j(t) - \phi_k(t)$ is constant^d, regardless of the type of oscillator.

2.4 Effect of linear mixing on synchrony

Assume that we have N sources which have PLFs of 1 with each other. Let $\mathbf{s}(t)$, for $t = 1, \dots, T$, denote the vector of sources and $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$ denote the mixed signals, where \mathbf{A} is the mixing matrix, which is assumed to be square and non-singular^e. Our goal is to find a square unmixing matrix \mathbf{W} such that the estimated sources $\mathbf{y}(t) = \mathbf{W}^T \mathbf{x}(t) = \mathbf{W}^T \mathbf{A}\mathbf{s}(t)$ are as close to the true sources as possible, up to permutation, scaling, and sign change.

The effect of linear mixing on the PLF matrix is illustrated in Figure 1 for a set of simulated sources. This set has three sources, with PLFs of 1 with each other. These sources are of the form (1) with negligible phase jitter, and the phase lags α_j are $0, \frac{\pi}{6}$, and $\frac{\pi}{3}$ radians, respectively. The common oscillation is a time-dependent sinusoid. The amplitudes are generated by adding a small constant baseline to a random number of "bursts" with Gaussian shape. Each "burst" has a random center and a random width, and each source amplitude has 1 to 5 such "bursts".

The first row of Figure 1 shows on the left the original sources and on the right their PLF matrix. The second row depicts the mixed signals $\mathbf{x}(t)$ on the left and their PLFs on the right; the mixing matrix has random entries uniformly distributed between -1 and 1 . It is clear that the mixed signals have lower pairwise PLFs than the sources, although signals 2 and 3 still exhibit a rather high mutual PLF. This example suggests that linear mixing of synchronous sources reduces their synchrony, a fact that will be proved in Section 3.3, ahead; this fact will be used to extract the sources from the mixtures by trying to maximize the PLF of the estimated sources.

3 Algorithm

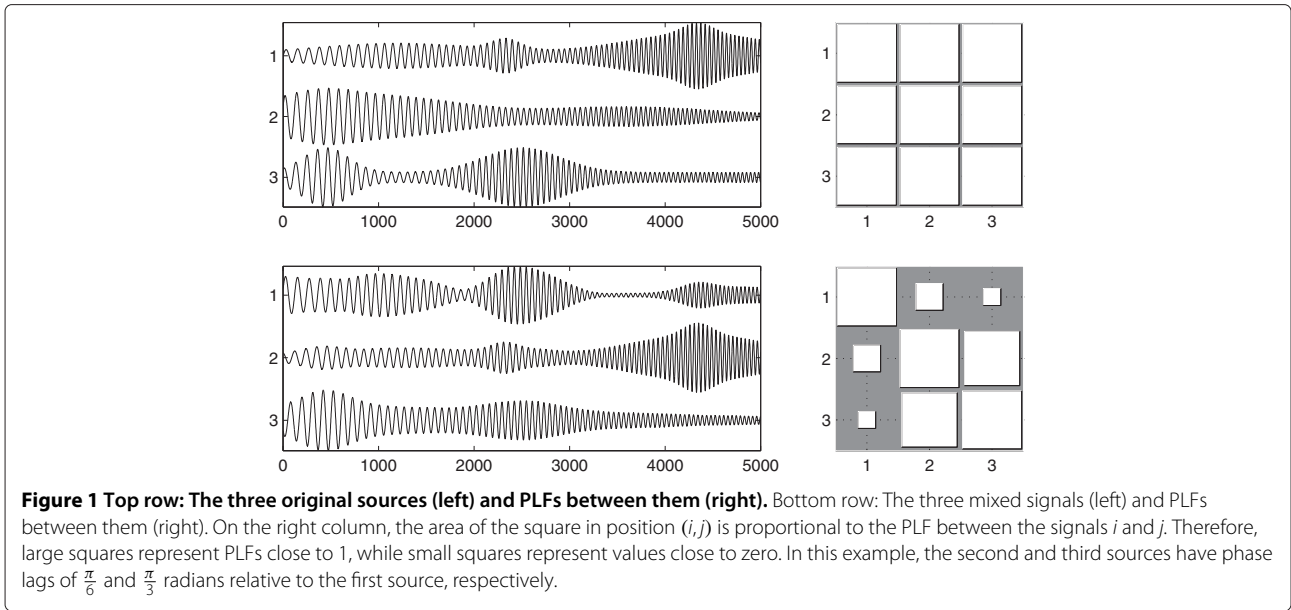
In this section, we describe the IPA algorithm. As mentioned in Section 1, this algorithm first performs subspace separation, and then performs separation within each subspace. In this article, we only study the performance of IPA in the case where all the sources are phase-locked; in this situation, the inter-subspace separation can entirely be skipped, since there is only one subspace of locked sources. Therefore, we will not discuss here the part of IPA relating to subspace separation; the reader is referred to [14] for a discussion on that subject.

3.1 Preprocessing

3.1.1 Whitening

As happens in ICA and other source separation techniques, whitening is a useful preprocessing step for IPA. Whitening, or sphering, is a procedure that linearly transforms the data so that the transformed data have the identity as its covariance matrix; in particular, the whitened data are uncorrelated [7]. In ICA, there are clear reasons to pursue uncorrelatedness: independent data are also uncorrelated, and therefore whitening the data already fulfills one of the required conditions to find independent sources. If \mathbf{D} denotes the diagonal matrix containing the eigenvalues of the covariance matrix of the data and \mathbf{V} denotes an orthonormal matrix which has, in its columns, the corresponding eigenvectors, then whitening can be performed in a PCA-like manner by multiplying the data $\mathbf{x}(t)$ by a matrix \mathbf{B} , where [7]

$$\mathbf{B} = \mathbf{D}^{-1/2} \mathbf{V}^T. \quad (4)$$



The whitened data are given by $\mathbf{z}(t) = \mathbf{B}\mathbf{A}\mathbf{s}(t)$. Therefore, whitening merely transforms the original source separation problem with mixing matrix \mathbf{A} into a new problem with mixing matrix $\mathbf{B}\mathbf{A}$. The advantage is that $\mathbf{B}\mathbf{A}$ is an orthogonal mixing matrix, and its estimation becomes easier [7].

The above reasoning is not valid for the separation of phase-locked sources. However, under rather general assumptions, satisfied by the data studied here, it can be shown that whitening places a relatively low upper bound on the condition number of the equivalent mixing matrix (see [28] and references therein). Therefore, we always whiten the mixture data before applying the procedures described in Section 3.2.

3.1.2 Number of sources

As will be seen below, IPA assumes knowledge of the number of sources, and also assumes that the mixing matrix is square: if this is not the case, a simple procedure can be used to detect the number of sources and to transform the data to obey these constraints. If the mixing process is noiseless and is given by $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$, where \mathbf{A} has more rows than columns and has maximum rank^f, the number of non-zero eigenvalues of the covariance matrix of \mathbf{x} is N , where N is the number of sources (or equivalently, the number of columns of \mathbf{A}). If the mixture is noisy with a low level of i.i.d. Gaussian additive noise, the former zero-valued eigenvalues now have small non-zero values, but detection of N is still easy to do by detecting how many eigenvalues are large relative to the plateau level of the small eigenvalues [7]^g. After N is known, the data need only be multiplied by a matrix $\mathbf{B}' = \mathbf{D}'^{-1/2}\mathbf{V}'^T$ in a similar fashion to Equation (4), where \mathbf{D}' is a smaller $N \times N$

diagonal matrix containing only the N largest eigenvalues in \mathbf{D} and \mathbf{V}' is a rectangular matrix containing only the N columns of \mathbf{V} corresponding to those eigenvalues. The mixture to be separated now becomes

$$\mathbf{x}'(t) = \mathbf{B}'\mathbf{x}(t) = \mathbf{B}'\mathbf{A}\mathbf{s}(t). \quad (5)$$

Since $\mathbf{B}'\mathbf{A}$ is a square matrix and the number of sources is now given simply by the number of components of \mathbf{x}' , the problem now has a known number of sources and a square mixing matrix.

A remark should be made about complex-valued data. The above procedure is appropriate when both the mixing matrix and the sources are real-valued. If both the mixing matrix and the sources are complex-valued, Equation (4) still applies (\mathbf{V} will now have complex values). However, in our case the sources and measurements are complex-valued (due of the Hilbert transform), but the mixing matrix is real. When this is the case, Equation (4) is not directly applicable. The above procedure must instead be applied not to the original data $\mathbf{x}(t)$, but to new data \mathbf{x}_0 with twice as many time samples, given by $\mathbf{x}_0(t) = \mathcal{R}(\mathbf{x}(t))$ for $t = 1, \dots, T$ and $\mathbf{x}_0(t) = \mathcal{I}(\mathbf{x}(t - T))$ for $t = T + 1, \dots, 2T$, where \mathcal{R} and \mathcal{I} denote the real and imaginary parts of a complex number, respectively. The matrix \mathbf{B} which results from applying Equation (4) to \mathbf{x}_0 (or \mathbf{B}' if appropriate) is then applied to the original data \mathbf{x} as before, and the remainder of the procedure is similar [28].

3.2 Separation of phase-locked sources

The goal of the IPA algorithm is to separate a set of N fully phase-locked sources which have linearly been mixed. Since these sources have a maximal PLF with each

other and the mixture components do not (as motivated in Section 2.4 above and proved in Section 3.3 below), we can unmix them by searching for projections that maximize the resulting PLFs. Specifically, this corresponds to finding a $N \times N$ matrix \mathbf{W} such that the estimated sources, $\mathbf{y}(t) = \mathbf{W}^T \mathbf{x}(t) = \mathbf{W}^T \mathbf{A} \mathbf{s}(t)$, have the highest possible PLFs.

The optimization problem that we shall solve is

$$\max_{\mathbf{W}} (1 - \lambda) \sum_{\substack{j,k=1 \\ j>k}}^N \varrho_{jk}^2 + \lambda \log |\det \mathbf{W}| \quad (6)$$

$$\text{s.t. } \|\mathbf{w}_j\| = 1, \text{ for } j = 1, \dots, N$$

where \mathbf{w}_j is the j th column of \mathbf{W} . In the first term, we sum the squared PLFs between all pairs of sources. The second term penalizes unmixing matrices that are close to singular, and λ is a parameter controlling the relative weights of the two terms. This second term serves the purpose of preventing the algorithm from finding, e.g., solutions where two columns j and k of \mathbf{W} are colinear, which trivially yields $\varrho_{jk} = 1$ (a similar term is used in some ICA algorithms [7]). Each column of \mathbf{W} is constrained to have unit norm to prevent trivial decreases of that term.

The optimization problem in Equation (6) is highly non-convex: the objective function is a sum of two terms, each of which is non-convex in the variable \mathbf{W} . Furthermore, the unit norm constraint is also non-convex. Despite this, as we show below in Section 3.3, it is possible to characterize all the global maxima of this problem for the case $\lambda = 0$ and to devise an optimization strategy taking advantage of that result.

The above optimization problem can be tackled through various maximization algorithms. Our choice was to use a gradient ascent algorithm with momentum and adaptive step sizes; after this gradient algorithm has run for 200 iterations, we use the BFGS algorithm implemented in MATLAB to improve the solution. The result of this optimization for the sources shown in Figure 1 is shown in Figure 2 for $\lambda = 0.1$, illustrating that IPA successfully recovers the original sources for this dataset.

3.3 Unicity of solution

In [14], we proved that a few mild assumptions on the sources, which are satisfied in the vast majority of real-world situations, suffice for a useful characterization of the global maxima of Problem (6): it turns out that there are infinitely many such maxima, and that they correspond either to correct solutions (i.e., the original sources up to permutation, scaling, and sign changes) or to singular matrices \mathbf{W} . More specifically, we proved the following: assume that we have a set of complex-valued and linearly independent sources denoted by $\mathbf{s}(t)$, which have a PLF of 1 with one another. Consider also linear combinations

of the sources of the form $\mathbf{y}(t) = \mathbf{C} \mathbf{s}(t)$ where \mathbf{C} is a square matrix of appropriate dimensions. Further assume that the following conditions hold:

1. Neither $s_j(t)$ nor $y_j(t)$ can identically be zero, for all j .
2. \mathbf{C} is non-singular.
3. The phase lag between any two sources is different from 0 or π .
4. The amplitudes of the sources, $a_j(t) = |s_j(t)|$, are linearly independent.

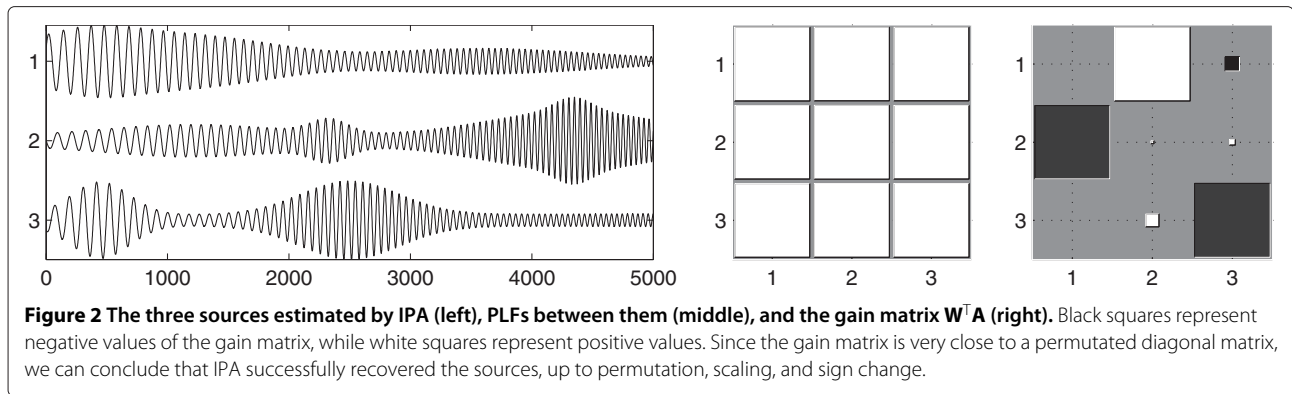
Then, the only linear combination $\mathbf{y}(t) = \mathbf{C} \mathbf{s}(t)$ of the sources $\mathbf{s}(t)$ in which the PLF between any two components of \mathbf{y} is 1 is $\mathbf{y}(t) = \mathbf{s}(t)$, up to permutation, scaling, and sign changes [14].

3.4 Comparison to ICA

The above result is simple, but some relevant remarks should be made. If the optimum is found using $\lambda = 0$ and the second assumption is not violated (or equivalently, $\det(\mathbf{C}) = \det(\mathbf{W}) \det(\mathbf{A}) \neq 0$, which is equivalent to $\det(\mathbf{W}) \neq 0$ if \mathbf{A} is non-singular), then we can be certain that the correct solution has been found. However, if the optimization is made using $\lambda = 0$, there is a possibility that the algorithm will estimate a bad solution where, for example, some of the estimated sources are all equal to one another (in which case the PLFs between those estimated sources is trivially equal to 1). On the other hand, if we use $\lambda \neq 0$ to guarantee that \mathbf{W} is non-singular, the unicity result stated above cannot be applied to the complete objective function. We call “non-singular solutions” and “singular solutions” those in which $\det(\mathbf{W}) \neq 0$ and $\det(\mathbf{W}) = 0$, respectively. The result expressed in Section 3.3 is thus equivalent to stating that “all non-singular global optima of Equation (6) with $\lambda = 0$ correspond to correct solutions”.

This contrasts strongly with ICA, where singular solutions are not an issue, because ICA algorithms attempt to find independent sources and one signal is never independent from itself [7]. In other words, singular solutions always yield poor values of the objective function of ICA algorithms. Here we are attempting to estimate phase-locked sources, and any signal is perfectly phase-locked with itself. Thus, one must always use $\lambda \neq 0$ in the objective function of Equation (6) when attempting to separate phase-locked sources.

We use a simple strategy to deal with this problem. We start by optimizing Equation (6) for a relatively large value of λ ($\lambda = 0.4$), and once convergence has been obtained, we use the result as the starting point for a new optimization, this time with $\lambda = 0.2$. The same process is repeated with the value of λ halved each time, until five such epochs have been run. The early optimization steps move the algorithm away from the singular solutions



discussed above, whereas the final steps are done with a very low value of λ , where the above unicity conditions are approximately valid. As the following experimental results show, this strategy can successfully prevent singular solutions from being found, while making the influence of the second term of Equation (6) on the final result negligible.

4 Experimental results

4.1 Data generation

As mentioned earlier, the main goal of this study is to study the applicability of IPA to real-world electrophysiological data from human brain EEG and MEG. The choice of the data for this study was not trivial, since we need to know the true sources in order to quantitatively measure the quality of the results. On the one hand, to know the actual sources in the brain would require simultaneous data from outside the scalp (EEG or MEG, which would be the mixed signals) and from inside the scalp (intra-cranial recordings, corresponding to the sources). If intra-cranial recordings are not available, results cannot quantitatively be assessed; they can only qualitatively be assessed by experts who can tell whether the extracted sources are meaningful or not. On the other hand, due to their extreme simplicity, synthetic data such as those used so far to illustrate IPA, shown in Figure 1, cannot be used to assess the usefulness of the method in real-world situations.

In an attempt to obtain “the best of both worlds”, we have generated a pseudo-real dataset from actual MEG recordings. By doing this, we know the true sources and the true mixing matrix, while still using sources that are of a nature similar to what one observes in real-world MEG. We begin by describing the process that we used to generate a perfectly phase-locked dataset; we then explain how we modified these data to analyze non-perfect cases as well. It is important to stress that the generation process described below has no relation to the one used to generate the data of Figure 1, even though both processes generate sources with maximum PLF.

Our first step was to obtain a realistic mixing matrix. To do so, we used the well-known EEGIFT software package [29]. This package includes a real-world sample EEG dataset with 64 channels. Using all the default options of the software package, we extracted 20 independent components from the data of Subject 1 in that dataset. The results that was important for us, in this process, were not the independent components themselves (which were discarded), but rather the 64×20 mixing matrix. As discussed in Section 3.1, we have opted for using a square mixing matrix, with little loss of generality. Therefore, we selected N random rows and N random columns of that mixing matrix (without repetition), and formed an $N \times N$ mixing matrix from the corresponding values of the original 64×20 matrix. We will later show results for datasets ranging from $N = 2$ to $N = 5$ sources; in the following, assume, for the sake of concreteness, that $N = 4$.

Having generated a physiologically plausible mixing matrix, the next step was to generate a set of four sources. For this, we used the MEG dataset studied previously in [30]^h, which has 122 channels with 17,730 samples per channel. The sampling frequency is 297 Hz, and the data have already been subjected to low-pass filtering with cut-off at 90 Hz. Since band-pass filtering is a very common preprocessing step in the analysis of MEG data [20-22] and is useful for the use of the Hilbert transform, we performed a further band-pass filtering with no phase distortion, keeping only the 18–24 Hz bandⁱ. The resulting filtered data were used to generate a complex signal through the Hilbert transform; these data were whitened as described in Section 3.1, and from the whitened data we extracted the time-dependent amplitudes and phases.

We then selected four random channels of these filtered MEG data. Since none of these MEG recordings were actually phase-locked (recall that they were themselves the result of a mixing process) and we wanted to study the performance on fully phase-locked sources (possibly corrupted by jitter, as explained below), we replaced the phase of the second of these channels with the phase of the first

channel with a constant phase lag of $\frac{\pi}{6}$ radians. The phase of the third channel was replaced with the phase of the first channel with a constant phase lag of $\frac{\pi}{3}$ radians, and that of the fourth channel with the phase of the first channel with a lag of $\frac{\pi}{2}$ radians. The amplitudes of the four sources were kept as the original amplitudes of the four random channels themselves. The process is illustrated in Figure 3. The above process, including the choice of the 4×4 submatrix, was repeated 100 times, with different initializations of the random number generator. This way of constructing the data ensured that the sources were fully phase-locked.

We also constructed datasets in which the sources were not perfectly phase locked. For this, we used the same 100 sets of sources, but with those sources now corrupted by phase jitter: each sample t of each source j was multiplied by $e^{i\delta_j(t)}$, where the phase jitter $\delta_j(t)$ was drawn from a random Gaussian distribution with zero mean and standard deviation σ . We tested IPA for σ from 0 to 20 degrees, in 5 degrees steps. One example with $\sigma = 5$ degrees is shown in Figure 4, and one with $\sigma = 20$ degrees is shown in Figure 5.

Finally, we studied the effect of N on the results of the proposed algorithm. We created 100 datasets similar to the jitterless datasets mentioned earlier, using $N = 2, 3$ and 5. In all of these, and similarly to the data with $N = 4$, we used sources with phase lags multiple of $\frac{\pi}{6}$.

4.2 Results

We measured the separation quality using two measures: the Amari performance index (API) [31] and the well-known signal-to-noise ratio (SNR). The API measures how far the gain matrix $\mathbf{W}^T \mathbf{A}$ is from a permuted diagonal matrix; the SNR measures how far the estimated sources are from the true sources. In summary, the API measures the quality of the estimation of the mixing matrix, while the SNR measures the quality of the estimation of the sources themselves.

Figure 6 presents the means and standard deviations of these measures for the 100 runs mentioned in Section 4.1, for each of the jitter levels. The results indicate that IPA

has very good performance on the jitterless case, in data of this kind, and that this level of performance is approximately maintained even in the presence of low levels of phase jitter, up to 5 degrees of standard deviation. Some deterioration in performance occurs from 5 to 10 degrees of phase jitter standard deviation, but with a SNR of 27 dB and an API below 0.1 the sources can still be considered to be well estimated.

The results for high jitter levels (σ equal to 15 or 20 degrees) show that there is a limit to IPA's robustness; this limit lies somewhere between 10 and 15 degrees. Equivalently, in terms of the PLF, the algorithm shows good robustness to PLF values smaller than 1 as long as they are above 0.95, but below that value its performance deteriorates progressively up to a PLF of approximately 0.9, at which point only partial separations are obtained.

Figure 7 shows the effect of varying the number of sources N . The figure shows that IPA can handle values of N up to $N = 5$ with only a slight decrease in performance.

Figure 7 also shows something rather peculiar: for $N = 2$, the results are mediocre (with an average API around 0.4)¹. This is not an effect of lowering the number of sources N , but rather an indirect effect of the phase lag between the sources. To verify this, we generated datasets of jitterless data with $N = 2$, using phase lags of $\frac{\pi}{12}$, $\frac{2\pi}{12}$ (the value used in Figure 7), $\frac{3\pi}{12}$, and $\frac{4\pi}{12}$ (100 datasets for each of these values). Figure 8 shows that a phase lag of $\frac{2\pi}{12}$ yields poor API values, as we already knew, but $\frac{3\pi}{12}$ yields very good values. Naively, one could conclude that when the sources have a phase lag of $\frac{2\pi}{12}$, or less, the separation cannot be accurately performed.

The effect is, however, not so clear-cut. The results for $N = 3, 4, 5$ also involve sources with phase lags of $\frac{\pi}{6}$, but the API values for those experiments are very good. We do not have a solid explanation for this fact; we conjecture that the presence of some pairs of sources with larger phase lags (e.g., for $N = 4$, the first and third sources have a phase lag of $\frac{\pi}{3}$ and the first and fourth sources have a phase lag of $\frac{\pi}{2}$) aids in the separation of all the sources.

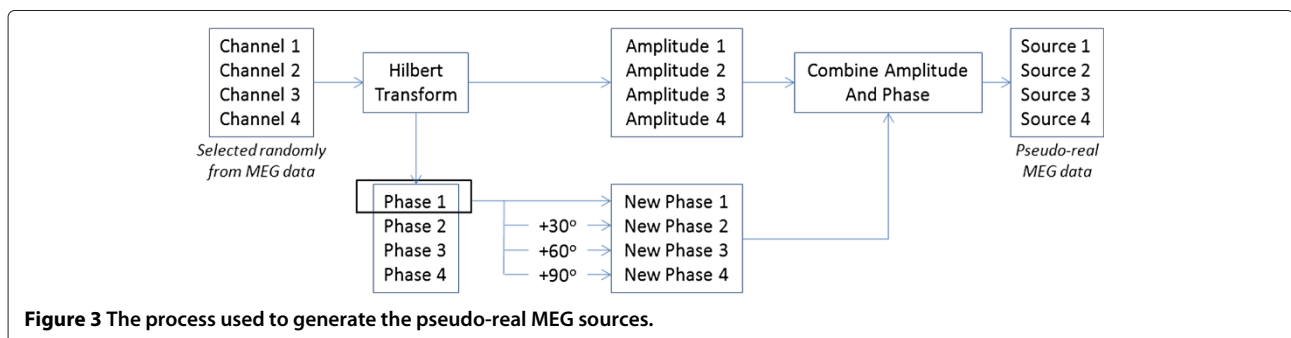
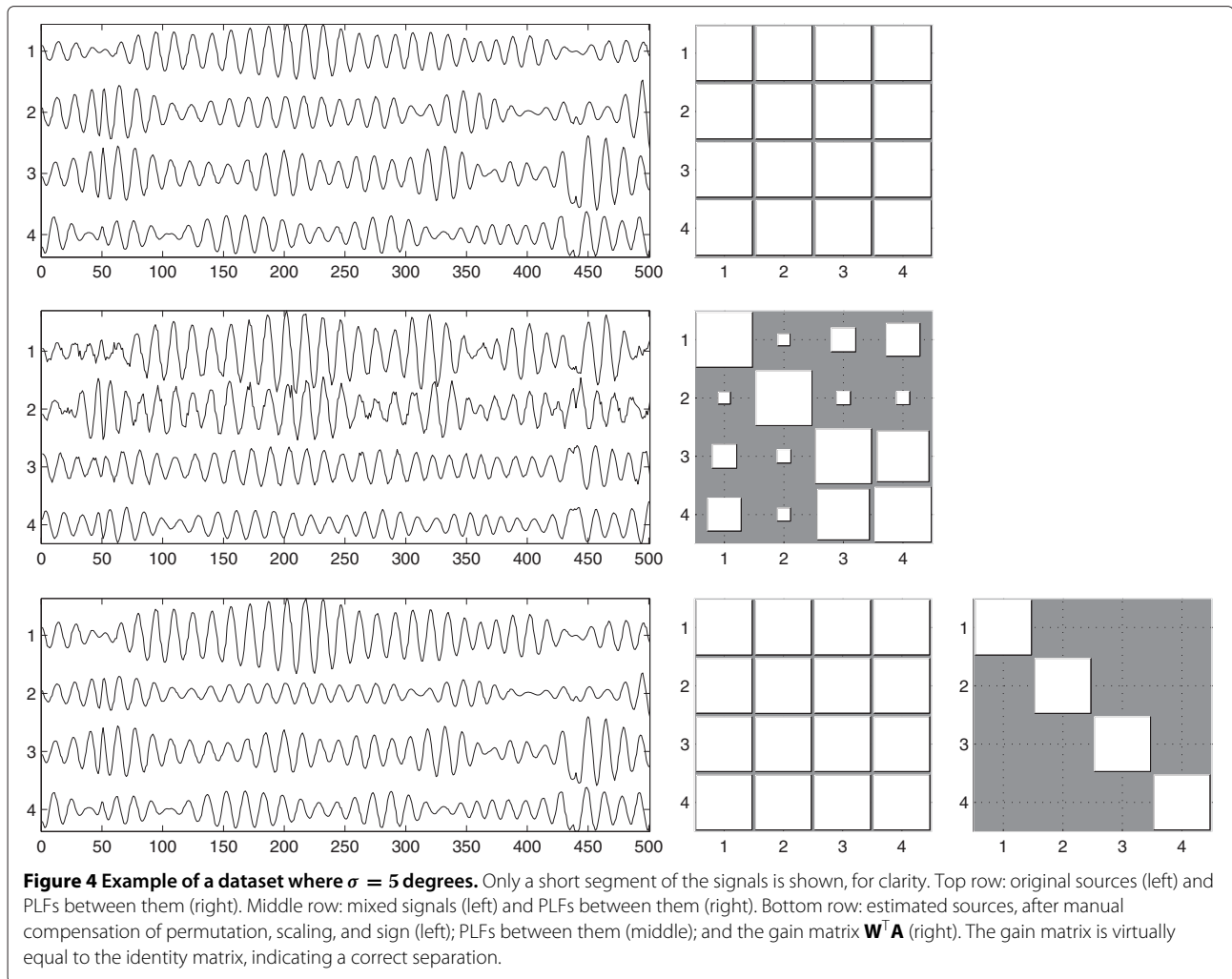


Figure 3 The process used to generate the pseudo-real MEG sources.



5 Discussion

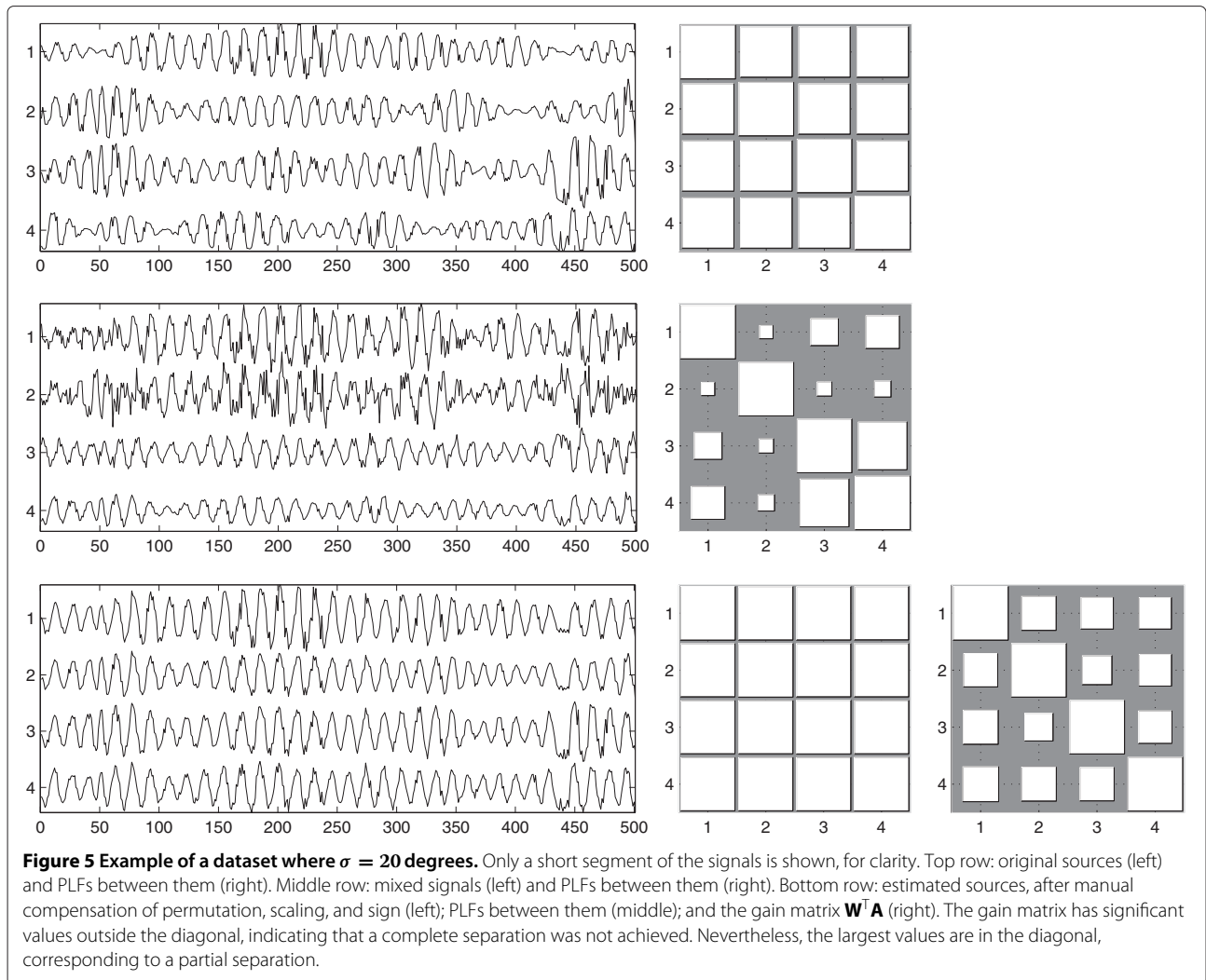
IPA has a parameter, λ , which controls the relative weights given to the optimization of the PLF matrix and to the penalization of close-to-singular solutions. Our optimization procedure starts with a high value of λ , which is lowered as the optimization progresses. We confirmed that this variation of the parameter's value is necessary: the quality of the results is noticeably degraded if λ is kept at a constant value, no matter how high or low it is. Table 1 confirms this: while $\lambda = 0.1$, the best fixed value, yields decent results, the results with a varying value of λ are considerably better. Furthermore, although the final epoch in the optimization is not done with $\lambda = 0$, we have verified that the results are virtually the same as if we had used $\lambda = 0$ at the last epoch.

The above paragraph illustrates something already mentioned in Section 3.4: separation of phase-locked sources is a non-trivial change from ICA because there are wrong, singular solutions that yield exactly the same values of the PLF matrix as the correct non-singular solutions. Our

approach to distinguish these two types of solutions consists in adding a term depending on the determinant of the matrix \mathbf{W} . This approach works correctly, as our results show. However, it is perhaps inelegant to do this through matrix \mathbf{W} , instead of doing it directly through the estimated sources. It would be preferable to replace this term with one depending directly on the estimated sources.

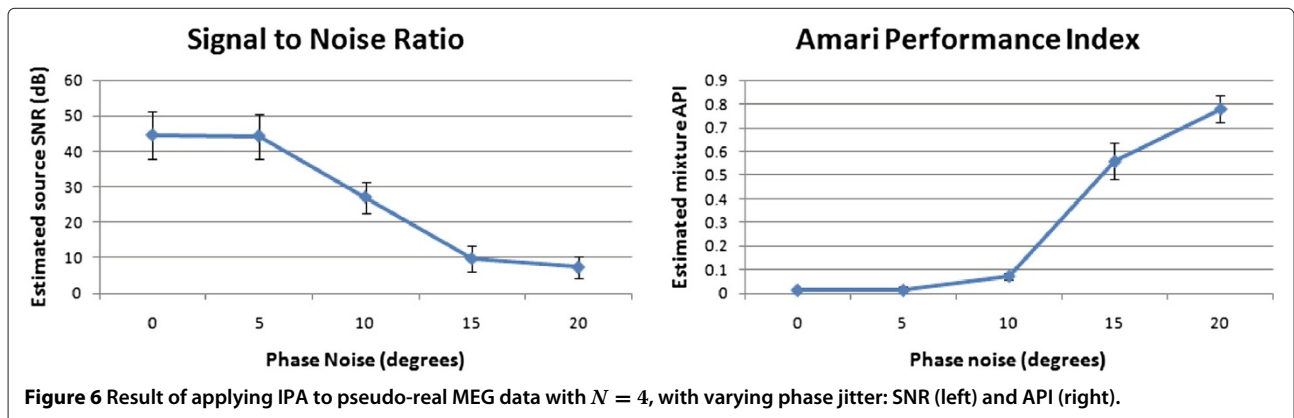
The size of the optimization variable, \mathbf{W} , is N^2 ; there are N constraints on this variable, yielding $N(N - 1)$ independent parameters. This means that the IPA algorithm is quadratic in the number of sources N , which is the main reason why we do not present results for $N > 5$; while running IPA on 100 datasets with $N = 2$ takes a few hours, doing so for $N = 5$ takes several days.

The results that we obtained show that IPA can separate perfectly locked MEG-like sources. However, while the phase locking in the jitterless pseudo-real MEG data is perfect, in real MEG data it will probably be less than perfect. This is the reason why we also studied data with phase jitter, which have pairwise PLFs smaller than 1. The



results indicate that IPA has some robustness to PLFs smaller than 1, but the sources still need to exhibit considerable phase locking for the separation to be accurate; weaker synchrony results only in partial separation. Note, however, that the partially separated data are usually still closer to the true sources than the original mixtures.

The comments made in the previous paragraph raise an additional optimization challenge: if the true sources have PLFs smaller than 1, optimization of the objective function in Equation (6) can lead to overfitting. The results presented here show that IPA has some robustness to sources which have a PLF smaller than 1, while



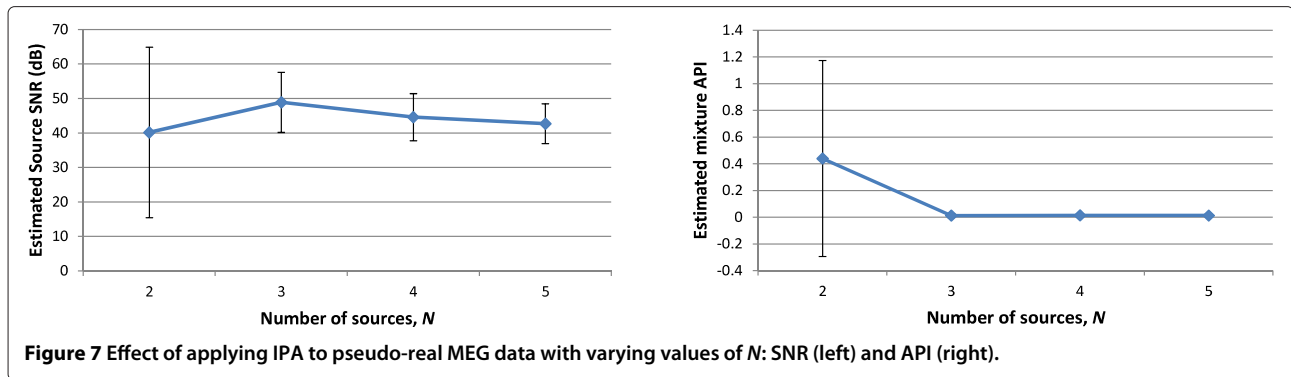


Figure 7 Effect of applying IPA to pseudo-real MEG data with varying values of N : SNR (left) and API (right).

being stationary (since the phase jitter is stationary, the distribution of the PLF does not vary with time). In real-world cases, it is likely that the PLF is non-stationary: for example, some sources may be phase-locked at the start of the observation period and not phase-locked at its end. While simple techniques such as windowing can be devised to tackle smaller time intervals where stationarity is (almost) verified, one would still need to find a way to integrate the information from different intervals. Such integration is out of the scope of this article.

One interesting extension of this article would be the separation of specific types of systems, such as van der Pol oscillators [27]. For those, fully entrained oscillators may even present a PLF < 1 , and a different measure of synchrony, tailored to those oscillators, may need to be used. Such a study would fall out of the scope of this article. Nevertheless, it is expected that additional knowledge of the oscillator type can be exploited to improve the algorithm's performance or its robustness to deviations from the ideal case.

One can derive a relationship between additive Gaussian noise (e.g., from the sensors) and the phase jitter used throughout this article. Figure 5 depicts, in the complex plane, a sample of a noiseless signal $x(t) \equiv$

$a(t)e^{i\phi(t)}$, to which complex noise $n(t)$ is added to form the noisy signal $x_n \equiv a(t)e^{i\phi(t)} + n(t)^k$. That figure also shows $n_{\perp}(t)$, which is the projection of $n(t)$ on the direction orthogonal to $x(t)$, and $x_{n\perp}(t) \equiv x(t) + n_{\perp}(t)$. Also depicted are $\phi(t)$, $\phi_n(t)$ and $\phi_{n\perp}(t)$, which are defined as the phases of $x(t)$, $x_n(t)$ and $x_{n\perp}(t)$, respectively.

It can easily be shown that, if $|n(t)| \ll |x(t)| = a(t)$, then $\phi_n(t) \approx \phi_{n\perp}(t) \approx \phi(t) + \frac{n_{\perp}(t)}{a(t)}$ [32]. This is an important relationship, because it shows that, under additive noise, portions of the signal with a large amplitude will have a better phase estimate than portions with a small amplitude, in which even small amounts of additive noise can severely disrupt the phase estimation. We thus believe that the PLF quantity, while attractive and elegant in theory, and despite working well with low amounts of additive noise [14], will probably need to be changed to factor in the amplitude in an appropriate way to deal with applications where considerable amounts of additive noise are present.

6 Conclusion

We have shown that IPA can successfully separate phase-locked sources from linear mixtures in pseudo-real MEG data. We showed that IPA tolerates deviations from the

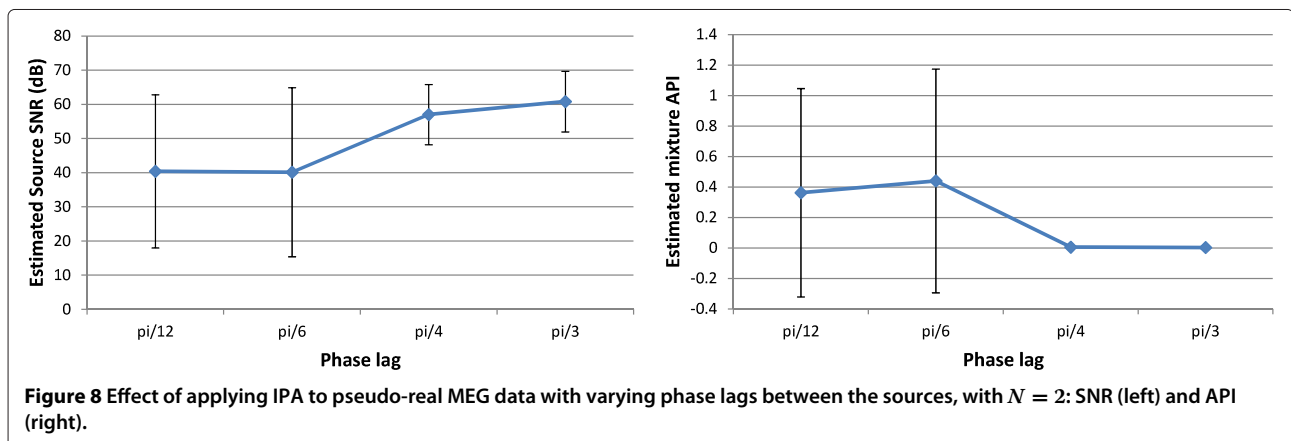


Figure 8 Effect of applying IPA to pseudo-real MEG data with varying phase lags between the sources, with $N = 2$: SNR (left) and API (right).

Table 1 Values of SNR and API for jitterless data with $N = 3$, for various fixed values of λ , as well as for the varying-lambda strategy detailed in the text

	λ	0.025	0.05	0.1	0.2	0.4
SNR	Fixed	17.5 ± 21.2	27.5 ± 18.0	34.4 ± 4.3	27.2 ± 3.6	13.5 ± 5.5
	Varying			48.9 ± 8.7		
API	Fixed	0.795 ± 0.570	0.369 ± 0.465	0.048 ± 0.057	0.079 ± 0.027	0.327 ± 0.097
	Varying			0.013 ± 0.015		

While the best fixed value, $\lambda = 0.1$, yields decent results, the results using a varying value of λ are consistently better, with a large margin.

ideal case, yielding excellent results for low amounts of phase jitter, and that it exhibits some robustness to moderate amounts of phase jitter. We also showed that it can handle numbers of sources up to $N = 5$. We believe that these results bring us closer to the goal of successfully separating phase-locked sources in real-world signals.

Endnotes

^aIn EEG and MEG, the sources are not individual neurons, whose oscillations are too weak to be detected from outside the scalp. In these cases, the sources are populations of closely located neurons oscillating together.

^bThe term “real-valued” is used here to distinguish from other phase-based algorithms where a complex quantity is used [14].

^cTechnically, this condition could be violated in a set with zero measure. Since we will deal with a discrete and finite number of time points, no such sets exist and this technicality is not important.

^dWe will also show results where this phase difference is not exactly constant; see Figure 6.

^eThese assumptions are not as restrictive as they may sound; see Section 3.1.

^fThis is usually called the over-determined case. The under-determined case, where \mathbf{A} has fewer rows than columns, is more difficult and is not addressed here.

^gThere are more rigorous criteria that can be used to choose N . Two very popular methods are the Akaike information criterion and the minimum description length. It is out of the scope of this article to discuss these two criteria; the reader is referred to [7] and references therein for more information.

^hFreely available from http://research.ics.tkk.fi/ica/eegmeg/MEG_data.html.

ⁱThe choice of this specific band is rather arbitrary. The band is narrow enough that the Hilbert transform will allow correct estimation of instantaneous amplitude and phase, but wide enough that the instantaneous frequency of the signals retains some variability. The passband is also of a similar width as in typical studies using MEG [20].

^jIt might appear contradictory that the average SNR has a good value, 40 dB, when the average API has a mediocre score. In reality, when the standard deviation of the SNR is very high, it is usually an indication that the separation is poor. As an example, consider a case where one source is very well estimated, with an SNR of 80 dB, and one is poorly estimated, with an SNR of 0 dB. The average SNR would be 40, but with a very high standard-deviation. Good values of the average SNR are indicators of a good separation only when the standard-deviation of the SNR is small.

^kIn most real applications, one will be dealing with models consisting of real signals to which real-valued noise is added. However, the linearity of the Hilbert transform allows the same type of analysis for that case as for the case of complex signals with complex additive noise which is considered here.

Competing interests

The author declare that they have no competing interests.

Acknowledgements

This work was partially supported by project DECA-Bio of Instituto de Telecomunicacoes, PEst-OE/EEI/LA0008/2011.

Author details

¹Instituto de Telecomunicacoes, Lisbon, Portugal. ²Aalto University School of Science, Espoo, Finland.

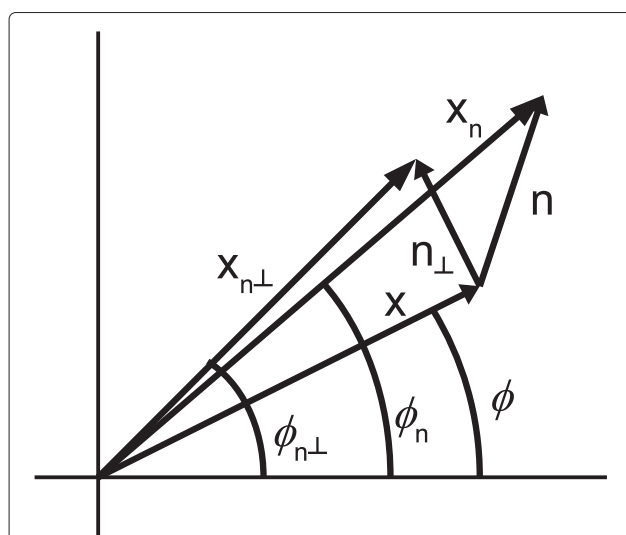


Figure 9 Diagram illustrating the relationship between phase jitter and additive noise. A single time sample is shown, and the time argument has been dropped for simplicity.

Received: 1 April 2012 Accepted: 2 February 2013
Published: 22 February 2013

References

1. M A Pikovsky, J Rosenblum, *Kurths, Synchronization: A Universal Concept in Nonlinear Sciences*. (Cambridge Nonlinear Science Series (Cambridge University Press), Cambridge, MA, 2001)
2. JM Palva, S Palva, K Kaila, Phase synchrony among neuronal oscillations in the human cortex. *J. Neurosci.* **25**(15), 3962–3972 (2005)
3. JM Schoffelen, R Oostenveld, P Fries, Imaging the human motor system's beta-band synchronization during isometric contraction. *NeuroImage.* **41**, 437–447 (2008)
4. PJ Uhlhaas, W Singer, Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology. *Neuron.* **52**, 155–168 (2006)
5. PL Nunez, R Srinivasan, AF Westdorp, RS Wijesinghe, DM Tucker, RB Silberstein, PJ Cadusch, EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalogr. Clin. Neurophysiol.* **103**, 499–515 (1997)
6. R Vigário, J Särelä, V Jousmäki, M Hämäläinen, E Oja, Independent component approach to the analysis of EEG and MEG recordings. *IEEE Trans. Biomed. Eng.* **47**(5), 589–593 (2000)
7. A Hyvärinen, J Karhunen, E Oja, *Independent Component Analysis*. (Wiley, New York, 2001)
8. M Akhtar, W Mitsuhashi, C James, Employing spatially constrained ICA and wavelet denoising for automatic removal of artifacts from multichannel EEG data. *Signal Process.* **92**, 401–416 (2012)
9. M de Vos, L de Lathauwer, S van Huffel, Spatially constrained ICA algorithm with an application in EEG processing. *Signal Process.* **91**, 1963–1972 (2011)
10. D Lee, H Seung, Algorithms for non-negative matrix factorization. *Adv. Neural Inf. Process. Syst.* **13**, 556–562 (2001)
11. TH Chan, WK Ma, CY Chi, Y Wang, A convex analysis framework for blind separation of non-negative sources. *IEEE Trans. Signal Process.* **56**, 5120–5134 (2008)
12. R de Frein, S Rickard, The synchronized short-time-Fourier-transform: properties and definitions for multichannel source separation. *IEEE Trans. Signal Process.* **59**, 91–103 (2011)
13. S Hosseini, Y Deville, H Saylani, Blind separation of linear instantaneous mixtures of non-stationary signals in the frequency domain. *Signal Process.* **89**, 819–830 (2009)
14. M Almeida, JH Schleimer, J Bioucas-Dias, R Vigário, Source separation and clustering of phase-locked subspaces. *IEEE Trans. Neural Netw.* **22**(9), 1419–1434 (2011)
15. M Almeida, J Bioucas-Dias, R Vigário, in *Proceedings of the International Conference on Independent Component Analysis and Signal Separation*, vol. 6365. Independent phase analysis: separating phase-locked subspaces, (2010), pp. 189–196
16. A Ziehe, KR Müller, in *International Conference on Artificial Neural Networks*. TDSEP—an efficient algorithm for blind separation using time structure, (1998), pp. 675–680
17. C Torrence, GP Compo, A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* **79**, 61–78 (1998)
18. AV Oppenheim, RW Schaffer, JR Buck, *Discrete-Time Signal Processing*. (Prentice-Hall International Editions, Englewood Cliffs, NJ, 1999)
19. MLV Quiyen, J Foucher, JP Lachaux, E Rodriguez, A Lutz, J Martinerie, FJ Varela, Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony. *J. Neurosci. Methods.* **111**, 83–98 (2001)
20. F Varela, JP Lachaux, E Rodriguez, J Martinerie, The Brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* **2**, 229–239 (2001)
21. E Niedermeyer, FHL da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. (Lippincott Williams and Wilkins, Philadelphia, 2005)
22. P Nunez, R Srinivasan, *Electric Fields of the Brain: the Neurophysics of EEG*. (Oxford University Press, New York, 2006)
23. B Gold, AV Oppenheim, CM Rader, in *Symposium on Computer Processing in Communications*. Theory and implementation of the discrete Hilbert transform, (1973)
24. M Breakspear, S Heitmann, A Daffertshofer, Generative models of cortical oscillations: neurobiological implications of the Kuramoto model. *Front. Human Neurosci.* **4**, 190–202 (2010)
25. Y Kuramoto, *Chemical Oscillations, Waves and Turbulences*. (Springer, Berlin, 1984)
26. S Strogatz, *Nonlinear Dynamics and Chaos*. (Westview Press, Boulder, 2000)
27. E Izhikevich, *Dynamic Systems in Neuroscience*. (MIT Press, Cambridge, MA, 2007)
28. M Almeida, R Vigário, J Bioucas-Dias, in *Proceedings of the International Conference on Latent Variable Analysis and Signal Separation*, vol. 7191. The role of whitening for separation of synchronous sources, (2012), pp. 139–146
29. T Eichele, S Rachakonda, B Brakedal, R Eikeland, VD Calhoun, EEGIFT: group independent component analysis for event-related EEG data. *Comput. Intell. Neurosci.* **2011**, 1–9 (2011)
30. R Vigário, V Jousmäki, M Hämäläinen, R Hari, E Oja, in *Advances in NIPS*. Independent component analysis for identification of artifacts in magnetoencephalographic recordings, (1997)
31. S Amari, A Cichocki, HH Yang, in *Advances in NIPS*, vol. 8. A new learning algorithm for blind signal separation, (1996), pp. 757–763
32. A Carlson, P Crilly, J Rutledge. *Communication Systems: An Introduction to Signals and Noise in Electrical Communication* (McGraw-Hill, New York, 2001)

doi:10.1186/1687-6180-2013-32

Cite this article as: Almeida et al.: Separation of phase-locked sources in pseudo-real MEG data. *EURASIP Journal on Advances in Signal Processing* 2013 **2013**:32.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com