Clinical Neurophysiology 119 (2008) 2677–2686

Contents lists available at ScienceDirect

Clinical Neurophysiology

journal homepage: www.elsevier.com/locate/clinph

Invited review

On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics

Marco Congedo *, Cédric Gouy-Pailler, Christian Jutten

GIPSA-lab (Grenoble Image Parole Signaux Automatique) CNRS (Centre National de la Recherche Scientifique) - Université Joseph Fourier - Université Stendhal - Grenoble INP (Institut National Polytechnique 961 rue de la Houille Blanche Domaine Universitaire-BP 46-38402, Grenoble, France

ARTICLE INFO

Article history:
Accepted 3 September 2008
Available online 6 November 2008

Keywords:
Blind source separation (BSS)
Independent component analysis (ICA)
Approximate joint diagonalization (AJD)
Electroencephalography (EEG)
Volume conduction
Fourier cospectra

ABSTRACT

Over the last ten years blind source separation (BSS) has become a prominent processing tool in the study of human electroencephalography (EEG). Without relying on head modeling BSS aims at estimating both the waveform and the scalp spatial pattern of the intracranial dipolar current responsible of the observed EEG. In this review we begin by placing the BSS linear instantaneous model of EEG within the framework of brain volume conduction theory. We then review the concept and current practice of BSS based on second-order statistics (SOS) and on higher-order statistics (HOS), the latter better known as independent component analysis (ICA). Using neurophysiological knowledge we consider the fitness of SOS-based and HOS-based methods for the extraction of spontaneous and induced EEG and their separation from extra-cranial artifacts. We then illustrate a general BSS scheme operating in the time-frequency domain using SOS only. The scheme readily extends to further data expansions in order to capture experimental source of variations as well. A simple and efficient implementation based on the approximate joint diagonalization of Fourier cospectral matrices is described (AJDC). We conclude discussing useful aspects of BSS analysis of EEG, including its assumptions and limitations.

1. Introduction

Recent studies on human electroencephalogram (EEG) are based on the theory of brain volume conduction. It is well established that the generators of brain electric fields recordable from the scalp are macroscopic post-synaptic potentials created by assemblies of pyramidal cells of the neocortex (Speckmann and Elger, 2005). Pyramidal cells are aligned and oriented perpendicularly to the cortical surface. Their synchrony is possible thanks to a dense net of local horizontal connections (mostly <1 mm). At recording distances larger than about three/four times the diameter of the synchronized assemblies the resulting potential behaves as if it were produced by electric dipoles; all higher terms of the multipole expansion vanish and we obtain the often invoked dipole approximation (Lopes da Silva and Van Rotterdam, 2005; Nunez and Srinivasan, 2006, Ch. 3). Three physical phenomena are important for the arguments we advocate in this study. First, unless dipoles are moving there is no appreciable delay in the scalp sensor measurement (Lopes da Silva and Van Rotterdam, 2005). Second, in brain electric fields there is no appreciable electro-magnetic coupling (magnetic induction) in the frequencies up to about 1 MHz, thus the quasi-static approximation of Maxwell equations holds throughout the spectrum of interest (Nunez and Srinivasan, 2006, p. 535-540). Finally, for source oscillations below 40 Hz it has been verified experimentally that capacitive effects are also negligible, implying that potential difference is in phase with the corresponding generator (Nunez and Srinivasan, 2006, p. 61). These phenomena strongly support the *superposition principle*, according to which the relation between neocortical dipolar fields and scalp potentials may be approximated by a system of linear equations (Sarvas, 1987). Whether this is a great simplification, we need to keep in mind that it does not hold true for all cerebral phenomena. Rather, it does at the macroscopic spatial scale we are interested in here.

A common approach to the study of human EEG is to describe patterns in space and time and link empirical findings with anatomical and physiological knowledge. The problem is characterized by high temporal resolution (about 1 ms) and low spatial resolution (several cm^3). For example, it has been estimated that without time averaging about 60 million contiguous neurons must be synchronously active as to produce observable scalp potentials (Nunez and Srinivasan, 2006, p. 21). Such a cluster would realistically extend over several cm^2 of cortical gyral surface, whereas disentangling fields emitted by cortical functional units may require much higher precision. Because of volume conduction, scalp EEG potentials describe a *mixture* of the fields emitted by several dipoles extending over large cortical areas. Practically, in order to

* Corresponding author. Tel.: +33 0 4 7657 4352; fax: +33 0 4 7657 4790.
E-mail address: Marco.Congedo@Gmail.com (M. Congedo).

1388-2457/$34.00
doi:10.1016/j.clinph.2008.09.007

improve the spatial resolution it is often necessary to trade in the temporal one operating some form of temporal averaging. In summary, the path followed by much of current EEG research is to "isolate" in space and time the generators of the observed EEG as much as possible, counteracting the mixing caused by volume conduction and maximizing the signal-to-noise ratio (SNR).

Over the years we have assisted to the development of several classes of methods to improve the spatial specificity. Those include, among others, surface and cortical Laplacian (Nunez and Srinivasan, 2006), equivalent dipole fitting (Mosher et al., 1992) and distributed minimum norm (model-driven) or minimum variance (data-driven) inverse solutions (Greenblatt et al., 2005; Lopes da Silva, 2004). Targeted attempts include sparsification approaches (Gorodnitsky et al., 1995; Cotter et al., 2005) and spatial filters known as beamformers (Rodrıguez-Rivera et al., 2006; Congedo, 2006). Surface Laplacian methods apply a spatial high-pass filtering to the scalp potential by estimating their second spatial derivative. They tend to overemphasize high spatial frequency and radial (to the scalp surface) dipolar fields. Inverse solutions seek *source localization* in a chosen solution space and rely on geometrical models of the head tissue. Unfortunately, the accurate description of EEG volume conduction is complicated by inhomogeneity (resistivity varies with type of tissue) and anisotropy (resistivity varies in different directions); therefore source localization methods are inevitably undermined by geometrical modeling error.

Another approach that persists in EEG literature is blind source separation (BSS). First studied in our laboratory during the first half of the 80's (Ans et al., 1985; Hérault and Jutten, 1986) BSS has enjoyed considerable interest worldwide only a decade later, inspired by the seminal papers of Jutten and Herault (1991), Comon (1994) and Bell and Sejnowski (1995). BSS has today greatly expanded encompassing a wide range of engineering applications such as speech enhancement, image processing, geophysical data analysis, wireless communication and biological signal analysis (Hyvärinen et al., 2001; Cichocki and Amari, 2002; Choi et al., 2005). Such ubiquity springs from the "blind" nature of the BSS problem formulation: no knowledge of volume conduction or of source waveform is assumed. The problem may be attacked from several perspectives; several hundred BSS algorithms have been proposed over the last 20 years with more added on every year. Typically, such methods are based on the cancellation of second order statistics (SOS) and/or of higher (than two) order statistics (HOS). Their commonality resides in the assumption of a certain degree of source *spatial independence*, which is precisely modeled by the cancellation of those statistics. Both HOS and SOS have been employed with success in EEG. They are today established for denoising/artifact rejection (Vigário, 1997; Jung et al., 2000; Vorobyov and Cichocki, 2002; Iriarte et al., 2003; Joyce et al., 2004; Kierkels et al., 2006; Fitzgibbon et al., 2007; Frank and Frishkoff, 2007; Halder et al., 2007; Phlypo et al., 2007; Romero et al., 2008; Crespo-Garcia et al., 2008), improving brain computer interfaces (Qin et al., 2004; Serby et al., 2005; Wang and James, 2007; Dat and Guan, 2007; Kachenoura et al., 2008) and for increasing the SNR of single-trial time-locked responses (Cao et al., 2002; Sander et al., 2005; Lemm et al., 2006; Tang et al., 2006; Guimaraes et al., 2007; Zeman et al., 2007). Yet, it appears that only four of the many existing algorithms have repeatedly occurred in EEG literature. They are known as FastICA (Hyvärinen, 1999), JADE (Cardoso and Souloumiac, 1993), InfoMax (Bell and Sejnowski, 1995) and SOBI (Belouchrani et al., 1997). FastICA, InfoMax and JADE are ICA (HOS) methods, while SOBI is a SOS method. JADE and SOBI are solved by *approximate joint diagonalization* (Cardoso and Souloumiac, 1993; Pham, 2001b; Yeredor, 2002; Ziehe et al., 2004; Vollgraf and Obermayer, 2006; Li and Zhang, 2007; Fadaili et al., 2007; Degerine and Kane, 2007), a powerful algebraic tool which allows promising extensions that we will consider in this study.

2. The BSS problem for the brain

For N scalp sensors and $M \leqslant N$ EEG dipolar fields with fixed location and orientation in the analyzed time interval, the linear BSS model simply states the superposition principle discussed above, i.e.,

$$v(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{\eta}(t), \tag{1.0}$$

where $\boldsymbol{v}(t) \in \mathrm{R}^N$ is the *sensor measurement vector*, $\boldsymbol{A} \in \mathrm{R}^{N \cdot M}$ is a time-invariant full column rank *mixing matrix*, $\boldsymbol{s}(t) \in \mathrm{R}^M$ holds the time-course of the source components and $\boldsymbol{\eta}(t) \in \mathrm{R}^N$ is additive noise, temporally white, possibly uncorrelated to $\boldsymbol{s}(t)$ and with spatially uncorrelated components. Our source estimation is given by

$$\hat{\boldsymbol{s}}(t) = \widehat{\boldsymbol{B}}v(t), \tag{1.1}$$

where $\boldsymbol{B} \in \mathrm{R}^{M \cdot N}$ is called the *demixing* or *separating matrix*. Hereafter the caret indicates a statistical estimation. Although this is the classical BSS model we need a few clarifications for the EEG case: first, by $\boldsymbol{\eta}(t)$ we model *instrumental* noise only. In the following we drop the $\boldsymbol{\eta}(t)$ term because the instrumental (and quantization) noise of modern EEG equipment is typically low (<1 μV). On the other hand, *biological* noise (extra-cerebral artifacts such as eye movements and facial muscle contractions) and *environmental* noise (external electromagnetic interference) may obey a mixing process as well, thus they are generally modeled as components of $\boldsymbol{s}(t)$, along with cerebral ones. Notice that while biological and environmental noise can be identified as separated components of $\boldsymbol{s}(t)$, hence removed, source estimation will be affected by the underlying cerebral *background noise* propagating with the same coefficients as the signal (Belouchrani and Amin, 1998). Second, the assumption of time-invariance of the mixing process in (1.0) must apply only locally. The demixing matrix is assumed fixed for a given temporal interval, but may be allowed to change (slowly) across successive intervals (Pham, 2001a; Li et al., 2006). Such a model allows changes in the location and orientation of dipole layers over time. The assumptions underlying model (1.0) are crucial for the success of the source separation, thus will be reconsidered in more details in the discussion.

3. A suitable class of solutions to the brain BSS problem

To tackle problem (1.1) assuming knowledge of sensor measurement only we need to reduce the number of admissible solutions. In this paper we are interested in weak restrictions converging toward condition

$$\hat{\boldsymbol{s}}(t) = \boldsymbol{G}\boldsymbol{s}(t), \tag{1.2}$$

where $\boldsymbol{s}(t)$ holds the time-course of the true (unknown) source processes and the *system matrix*

$$\boldsymbol{G} = \hat{\boldsymbol{B}}\boldsymbol{A} \approx \boldsymbol{\Lambda}\boldsymbol{P} \tag{1.3}$$

approximates a signed scaling (a diagonal matrix $\boldsymbol{\Lambda}$) and raw permutation ($\boldsymbol{P}$). Eq. (1.2) is obtained substituting (1.0) in (1.1) ignoring the noise term in the former. Whether condition (1.2) may be satisfied is a problem of *identifiability*, which establish the theoretical ground of BSS theory (Tong et al., 1990; Tong et al., 1991a,b; Tong et al., 1993; Cardoso, 1998a; Pham and Cardoso, 2001; Pham, 2002; Theis, 2004). In turn, matching condition (1.2) implies that we can recover faithfully the source *waveform* out of a *scale* (including sign) and *permutation* (order) indeterminacy. The idea suits EEG well, since the waveform bears meaningful physiological and clini-