Removal of BCG artifacts from EEG recordings inside the MR scanner: a comparison of methodological and validation-related aspects

Katrien Vanderperren[a,*], Maarten De Vos[a], Jennifer R. Ramautar[bc], Nikolay Novitskiy[b], Sara Assecondi[d], Bart Vanrumste[ae], Peter Stiers[f], Bea Van den Bergh[bg], Johan Wagemans[b], Lieven Lagae[c], Stefan Sunaert[h], Sabine Van Huffel[a]

[a]*Katholieke Universiteit Leuven, Department of Electrical Engineering, ESAT-SCD, Leuven, Belgium*

[b]*Katholieke Universiteit Leuven, Department of Psychology, Laboratory of Experimental Psychology, Leuven, Belgium*

[c]*Katholieke Universiteit Leuven, Department of Pediatric Neurology, Leuven, Belgium*

[d]*Ghent University, Department of Electronics and Information Systems, MEDISIP-IBBT-IbiTech, Ghent, Belgium*

[e]*Katholieke Hogeschool Kempen, Biosciences and Technology Department, Geel, Belgium*

[f]*Maastricht University, Faculty of Psychology and Neuroscience, Maastricht, The Netherlands*

[g]*Tilburg University, Department of Psychology, Tilburg, The Netherlands*

[h]*Katholieke Universiteit Leuven, Department of Radiology, Leuven, Belgium*

* Corresponding author: Katrien Vanderperren

Kasteelpark Arenberg 10 – bus 2446

B-3001 LEUVEN

BELGIUM

TEL: +32 16 321799 – FAX: +32 16 321970

katrien.vanderperren@esat.kuleuven.be

Abstract

Multimodal approaches are of growing interest in the study of neural processes. To this end much attention has been paid to the integration of electroencephalographic (EEG) and functional magnetic resonance imaging (fMRI) data because of their complementary properties. However, the simultaneous acquisition of both types of data causes serious artifacts in the EEG, with amplitudes that may be much larger than those of EEG signals themselves. The most challenging of these artifacts is the ballistocardiogram (BCG) artifact, caused by pulse-related electrode movements inside the magnetic field. Despite numerous efforts to find a suitable approach to remove this artifact, still a considerable discrepancy exists between current EEG-fMRI studies. This paper attempts to clarify several methodological issues regarding the different approaches with an extensive validation based on event-related potentials (ERPs). More specifically, Optimal Basis Set (OBS) and Independent Component Analysis (ICA) based methods were investigated. Their validation was not only performed with measures known from previous studies on the average ERPs, but most attention was focused on task-related measures, including their use on trial-to-trial information. These more detailed validation criteria enabled us to find a clearer distinction between the most widely used cleaning methods. Both OBS and ICA proved to be able to yield equally good results. However, ICA methods needed more parameter tuning, thereby making OBS more robust and easy to use. Moreover, applying OBS prior to ICA can optimize the data quality even more, but caution is recommended since the effect of the additional ICA step may be strongly subject-dependent.

Introduction

Simultaneous acquisition of electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) is a highly valuable method in the study of cerebral dynamics. The complementarity between the high temporal resolution of EEG and the high spatial resolution of fMRI can provide deeper insight into brain function and dysfunction (Babiloni et al., 2004). This advantage has already been exploited in numerous applications. For instance, the combination of EEG and fMRI allows localizing epileptic activity based on spike-triggered fMRI (Seeck et al., 2001; Krakow et al., 2001; Bénar et al., 2003; Lemieux et al., 2001). Other possible applications are the study of ongoing brain rhythms (Goldman et al., 2000; Goldman et al., 2002; Laufs et al., 2003; Moosmann et al., 2003) and cerebral activation during sleep (Czisch et al., 2002; Liebenthal et al., 2003; Schabus et al., 2007). Also the analysis of event-related brain responses based on multimodal information (Mulert et al., 2004; Debener et al., 2005; Debener et al., 2006; Eichele et al., 2008; Moosmann et al., 2008) becomes more and more popular. This latter application will benefit especially from a meaningful combined approach since event-related responses are low amplitude signals highly sensitive to experimental conditions and brain state.

An important disadvantage, however, is that EEG data are highly contaminated with two major artifacts when acquired in the MR scanner. Firstly, gradient artifacts occur due to switching magnetic fields during fMRI acquisition. These artifacts can have amplitudes 10 to 100 times larger than those of EEG signals. Several procedures have been proposed to remove this type of artifact from the EEG (e.g., Sijbers et al., 2000). Nevertheless,

since this artifact is invariant over time, most studies nowadays are consistent in using a subtracting procedure based on an average artifact template (Allen et al., 2000).

A bigger challenge is posed by the ballistocardiogram (BCG) artifact, produced by cardiac pulse-related movement of the scalp electrodes inside the magnetic field. Not only is the exact cause of this movement still a matter of investigation (e.g., Yan et al., 2009), the removal of this artifact is also a problematic issue, reported in many simultaneous EEG-fMRI studies (e.g., Debener et al., 2007). As a consequence, various methods have been suggested. Roughly, these methods can be subdivided into two main categories.

The first category of methods for BCG artifact removal is based on the channel-wise subtraction of an artifact template. The methods differ in the way this template is generated. One of the first methods to be introduced was similar to the procedure later adopted for gradient artifacts, employing a dynamic average artifact template (Allen et al., 1998). Subsequently, variations on this average template were developed based on median-filtering (Ellingson et al., 2004) and Gaussian weighted averaging (Goldman et al., 2000). More recently, it was suggested to use an Optimal Basis Set (OBS) of principal components for the template creation (Niazy et al., 2005). This technique relies on the idea that principal component analysis (PCA) applied to all artifact occurrences in each channel separately makes it possible to capture the temporal variations of the BCG artifact.

The second category of methods is based on blind source separation. There are several ways to perform this separation (see Methods section) and many have already been suggested for the removal of the BCG artifact. Assecondi et al. (2009), for instance, make

use of Canonical Correlation Analysis to perform blind source separation. The most widely reported blind source separation technique for BCG artifact removal, however, is Independent Component Analysis (ICA) (Bénar et al., 2003; Srivastava et al., 2005; Mantini et al., 2007). This is a signal processing technique that can be used to recover sources from a set of simultaneously recorded signals, assuming that they result from a linear mixing of independent source signals (Comon, 1994). ICA applied to EEG data contaminated with BCG artifacts can potentially identify both brain- and artifact-related sources, given that they are independent, thereby cleaning up the EEG by removing the artifactual sources.

As will be explained in more detail in the Methods section, most ICA algorithms assume stationarity of the sources. Since the BCG artifact shows a considerable spatial variation across its occurrence (Vanderperren et al., 2007), satisfying this assumption can be problematic. For this reason, it was suggested to apply OBS prior to ICA (Debener et al., 2005) instead of applying ICA directly on the EEG data. This approach would combine the strengths of both methods, removing the major part of the artifact with OBS and its residuals with ICA.

Despite the numerous approaches suggested, there is so far little consensus on what method gives the best data quality (Mantini et al., 2007; Debener et al., 2008). Grouiller et al. (2007) made an attempt to evaluate different methods by comparing their effect on continuous EEG data. However, since the study of event-related responses with simultaneous EEG-fMRI measurements is a growing field, the performance of the methods should also be investigated there. In Debener et al. (2007) the two most popular methods (OBS and ICA) and their combination (OBS-ICA) were compared on auditory

event-related potentials (ERPs). It was concluded that OBS and OBS-ICA could efficiently remove BCG artifacts and substantially improve the quality of the auditory responses. Although this study has led to a better understanding of the performance of OBS and ICA for BCG artifact removal, only little attention has been paid to the different method-dependent parameters to be selected and validation has only been performed with average measures. The present study aims, therefore, at furthering the selection of the most appropriate method by investigating these remaining uncertainties.

The aim of this paper is to clarify the confusion that arises due to the manifold of current approaches. Our first aim was, therefore, a direct comparison of the quality of artifact removal of the most widely used methods, namely OBS, ICA and OBS-ICA applied to ERP datasets. More specifically, data with ERPs from visual tasks were investigated, both in two *in vivo* datasets and in a carefully controlled simulation study. The application on visual ERPs has, to our knowledge, only got little attention so far. Furthermore, the combination of both *in vivo* and simulation data provides us with a more profound understanding of the performance of the BCG removal methods on realistic experimental data as well as a precisely controlled situation.

Secondly, we set out to establish for each of these methods the optimal parameter settings, yielding the best possible performance. This issue has been somewhat disregarded in previous studies, although an optimal choice of parameters is essential for the quality of the results. Thirdly, we set out to evaluate the effect of these optimized methods at all relevant levels of EEG datasets, ranging from average measures comparing the ERPs inside and outside the magnetic field to statistical power in discriminating experimental conditions in the ERPs. Hence, the six validation criteria included not only

account for a good artifact removal but also for an accurate recovery of the underlying data.

Materials and methods

In the following subsections first the various steps of the acquisition of *in vivo* experimental data are discussed after which the construction of simulated datasets is explained. Subsequently, the employed methods for BCG artifact removal are described, together with the applied validation criteria.

Subjects

Six healthy subjects (4 female and 2 male, aged 21-33) with no history of neurological or cardiological disorders participated in this study. Written informed consent was obtained in accordance with the local ethical committee guidelines. Experiments were in compliance with national legislation and the Code of Ethical Principles for Medical Research Involving Human Subjects of the World Medical Association. Data from one female subject were eventually discarded because of poor data quality due to severe body movement.

Tasks

Two visual tasks were presented to the participants. The corresponding stimuli were displayed with the Presentation 11.0 software (Neurobehavioral Systems, San Francisco, USA).

The first task was a visual detection paradigm with segments of circular black-and-white checkerboard stimuli presented one at a time in randomized sequences to one of the four quadrants of the visual field (Di Russo et al., 2002). In addition, a large circular black-and-white checkerboard was presented as a central stimulus on the middle part of the screen. Subjects were asked to press a button upon detection of each of these stimuli. Per run 100 stimuli (20 of each type) were shown to the participants. In each condition the stimulus was presented for 100 ms preceded by a fixation cross for 500 ms. The Inter-Stimulus Interval (ISI) varied between 900 ms and 2400 ms. This simple detection task was selected as it is known to evoke robust P1 and N1 components useful for validating the artifact removal. The P1 component is a positive deflection in the ERP around 100 ms after the stimulus onset. The N1 component is the negative deflection following this P1. The second task the subjects were requested to perform, is a cognitive control task. More specifically, a visual Go/NoGo paradigm was employed to investigate the mechanisms between generating and withholding responses (Falkenstein et al., 1999; Bokura et al., 2001; Nieuwenhuis et al., 2003). This task requires fast, simple responses to Go stimuli (red square) and inhibition of any responses to NoGo stimuli (red circle). Each run consisted of 100 trials of which 80 were Go trials. In each condition the stimulus was presented for 150 ms preceded by a fixation cross for 100 ms. The ISI varied between 850 ms and 5850 ms. The Go/NoGo task has led to the identification of ERP components such as the NoGo N2 and P3 that are strongly augmented on NoGo-trials, as compared with Go trials, particularly at frontocentral sites.

Data acquisition

All subjects performed 4 runs of each task, alternately. For validation purposes the EEG data were not only acquired in the static magnetic field of the scanner, but also outside the scanner room. Since this study focuses on BCG artifacts, the measurements were restricted to EEG data without simultaneous fMRI acquisition. The EEG data were collected from 62 standard scalp sites using the BrainAmp MR+ system (BrainProducts, Munich, Germany) with a sampling rate of 5 kHz. Two additional electrodes were placed below the left eye and on the right lower back to monitor eye blinks and electrocardiograms (ECGs), respectively. All 64 channels were recorded with Cz as reference and POz as ground. Electrode impedances were kept below 10 kΩ. For the measurements inside the scanner a Siemens 3T Allegra scanner (Siemens, Munich, Germany) was used.

Simulation study

To generate a set of simulated signals, a short fragment of continuous EEG data was acquired from one subject (male, aged 46), who did not perform any task while lying in the scanner. For practical reasons, these measurements were restricted to 5 minutes and conducted in a Philips 3T Intera scanner (Philips, Amsterdam, The Netherlands). This dataset can be assumed to contain background EEG contaminated by realistic BCG artifacts and instrumental noise.

Subsequently, Go and NoGo grand average ERPs obtained during outside measurements were added to this dataset. These grand average ERPs were retrieved from Go/NoGo data acquired outside the scanner on 3 new subjects (2 male, 1 female, aged 18-25) with the exact same protocol as described earlier, and were used as models for Go and NoGo

single trial responses. To extract the grand average ERPs, data were epoched on Go and NoGo stimuli and averaged after baseline removal and extreme value rejection.
The contribution of the background EEG and BCG artifacts with respect to the ERPs was amplified with a varying factor, to obtain different SNR values. The results of this simulation study were validated with the square root of the mean squared values of the differences between the original grand average ERPs and the ERPs obtained after BCG removal.

EEG preprocessing

The acquired EEG data were subjected to the following preprocessing steps.
Preprocessing as well as actual artifact removal were performed in the MATLAB 7.7.0 (R2008B) (The Mathworks Inc, Natick, Massachusetts, USA) environment with the EEGLAB 5.03 toolbox (Delorme et al., 2004).
First, all data were lowpass-filtered using a Finite Impulse Response (FIR) filter with a cut-off frequency of 30 Hz. Subsequently, the resulting data were downsampled to 250 Hz. The timing of the heartbeats was retrieved by detecting the QRS complexes on the ECG channel. This QRS detection, based on a modified Teager energy operator (Kim et al., 2004), was performed using the EEGLAB plug-in FMRIB version 1.2 (Niazy et al., 2005).
In the following paragraphs the actual methods for BCG removal are explained together with the parameters to set for each method. For a better understanding an overview is given in Figure 1.

OBS method

As stated in Niazy et al. (2005) OBS is based on the assumption that each BCG artifact occurrence in each channel is independent of any previous occurrence. Furthermore, following the same principle, all artifact occurrences are assumed to be sampled from an unknown set of possible variations. To capture these principal variations, OBS uses PCA, applied to all artifact occurrences per channel.

The first step of OBS (Niazy et al., 2005) consists in shifting all QRS complex timings forward in time by 210 ms. This time period is chosen because it is known as the standard delay between QRS complexes and the corresponding BCG artifact occurrences. Next, for each channel a matrix of BCG artifact occurrences is constructed and PCA is performed. The BCG artifact template is created based on a selected number of the resulting principal components, called the Optimal Basis Set. In order to obtain the cleaned data, the OBS per channel is fitted to, and subtracted from, each artifact occurrence.

In this study, the OBS method was performed as available in the EEGLAB plug-in FMRIB. One parameter needs to be selected, namely the number of principal components (NPC) used to create the artifact template. Most studies use the recommended value of 3 for this parameter, but, as one of the aims of this paper is to investigate the effect of each parameter, this number was varied between 1 and 7 (with a step size of 1).

ICA method

The EEG represents a complex combination of numerous sources of neural activity. Because EEG is acquired with a number of electrodes on different positions on the scalp,

each electrode will measure a weighted sum of these different sources. Mathematically, this can be represented with

$$X = A \cdot S + N \quad (1)$$

where X is the matrix of acquired signals $x_i(t)$ and S the matrix of source signals $s_i(t)$. The coefficients of A determine the contribution of the individual sources $s_i(t)$ in the measured signals and N represents the Gaussian noise. The aim of blind source separation comes down to the determination of the source signals $s_i(t)$ from the measured signals $x_i(t)$. These $s_i(t)$ can be estimated as

$$S = W \cdot X \quad (2)$$

with W the unmixing matrix. However, since both coefficients and sources are unknown, it is generally impossible to determine them without imposing additional constraints. Therefore, several possible assumptions about the sources have been proposed in order to obtain a unique decomposition. The most well-known is the constraint of statistical independence as imposed in ICA. As indicated in the introduction, ICA is a signal processing technique that recovers independent sources from a set of simultaneously recorded signals that result from a linear mixing of the source signals (Comon et al., 1994). In Srivastava et al. (2005) and Mantini et al. (2007) the ability of ICA to identify and remove BCG-related independent components has been shown.

During the application of ICA however, several choices need to be made. Not only several ICA algorithms exist, but also it is necessary to decide how many components should be retrieved. Once these components are obtained, a third important aspect is the approach followed to select the artifact-related components. In this study these three

mentioned choices were regarded as ICA parameters and will be discussed in more detail in the following sections.

1. Algorithm

In order to extract independent sources, various ICA algorithms can be used. In this study, four different implementations, namely Joint Approximate Diagonalization of Eigenmatrices (JADE), Second Order Blind Identification (SOBI), FastICA and extended Infomax ICA were compared. These algorithms were selected because they are well-known and most widely used in neuroscience applications (Stone, 2002; Makeig et al., 2004; Joyce et al., 2004).

JADE is one of the earliest ICA implementations and has a procedure based on higher-order statistics (Cardoso et al., 1993). It diagonalizes the fourth-order cumulant matrices of the estimated sources. In Urrestarazu et al. (2004), JADE proved to be effective for removing artifacts from EEG data.

SOBI (Belouchrani et al., 1993) uses second order statistics to decompose the measurements and is employed in numerous EEG studies (e.g., Tang et al., 2005). It is based on a joint diagonalization of correlation matrices. SOBI is most appropriate for sources that are individually correlated in time, but mutually uncorrelated. Furthermore, it exploits non-stationarity, which is typical for EEG data. In a recent study of Klemm et al. (2009), SOBI proved to be among the best methods for separating EEG data.

FastICA on the other hand, is described as a computationally efficient method for performing the estimation of ICA (Hyvärinen et al., 2001). It uses a so-called fixed-point iteration scheme that has been proven to be 10-100 times faster than conventional

gradient descent methods. The sources are extracted one-by-one based on their kurtosis. FastICA was for the first time used for the removal of the BCG artifact in Mantini et al. (2007).

Infomax ICA performs the separation in independent sources based on the information maximization principle (Bell and Sejnowski, 1995). More specifically the independent sources are retrieved by maximizing the entropy. Infomax is a very popular algorithm for EEG decomposition and has previously been used in the BCG removal studies of Srivastava et al. (2005) and Debener et al. (2007). In the current study the extended Infomax approach, as proposed by Lee et al. (1999), was applied.

To our knowledge, detailed comparisons of the effects of these different algorithms for BCG removal are so far lacking. In this paper, we therefore investigated the effect of selecting a particular algorithm in more detail, by considering the four abovementioned ICA implementations.

2. Number of components

In BSS problems it is so far not possible to determine beforehand the exact number of sources generating the acquired signals. However, since the number of channels in our experiments is reasonably high (62) compared to other simultaneous EEG-fMRI studies (Niazy et al., 2005; Mantini et al., 2007; Grouiller et al., 2007), it is very well possible that the real number of sources underlying the measured data does not equal the number of channels. As a consequence, reducing the dimensionality of the data before estimating the sources, might simplify the complexity of the problem for the actual iterative

algorithm. In previous studies on ICA for BCG removal, the number of components was always chosen equal to the number of channels.

To investigate the effect of dimensionality reduction, this study did not rely on a fixed choice for the number of components but explored several possibilities. The number of components was varied among the following choices: 15-20-25-30-35-40-50-62. This selected range and sampling allowed us to determine the effect of this parameter.

3. Selection criteria

Once ICA has been performed and the underlying sources have been retrieved, it is crucial to carefully identify and select the artifact-related components. The first attempts to use ICA for BCG removal were mainly based on manual selection of the artifactual components (e.g., Huiskamp et al., 2005). Subsequently, several automated selection criteria have been proposed in literature, the most important ones of which rely on correlation and variance properties (Srivastava et al., 2005; Mantini et al., 2007; Debener et al., 2008). In this paper we extended these selection criteria. In particular, we compared the accuracy of removal after selection based on correlation, frequency behavior, autocorrelation, variance and characteristics of reconstructed signals.

a) Correlation

The most widely reported selection criterion is based on the amount of correlation of the resulting independent components with the ECG channel (Srivastava et al., 2005; Mantini et al., 2007). The BCG artifact, however, does not resemble the ECG pattern at all; therefore another reference signal might be more effective. Srivastava et al. (2005)

introduced the idea to use a template for the BCG artifact as reference signal. For this purpose averaged BCG artifacts were concatenated per channel and subsequently averaged over all channels. However, since BCG artifacts are known to have an inverse polarity at both sides of the head (Debener et al., 2007; Vanderperren et al., 2007), it is possible that by using an average over all channels for the BCG template, a substantial amount of the artifactual activity will be averaged out. Therefore, the BCG template might be improved by squaring these concatenated averaged artifacts before taking the mean over the channels, in order to emphasize the BCG artifacts. Furthermore, in order to take some temporal variation into account, the averaging per channel should be performed in windows.

In the present study this latter procedure was compared with the two previously suggested ones. After correlating the ICA components with the suggested templates, a threshold was set on the correlation so as to identify all components with a higher correlation as artifact-related. For the actual threshold with all the three templates, it was chosen to start from 0.20, as this value is known from the literature (e.g., Mantini et al., 2007), and to include the surrounding values 0.15 and 0.30.

b) Frequency content

In Vanderperren et al. (2007), it has been shown that the presence of the BCG artifact is characterized by peaks in the frequency spectrum at the heart rate frequency and several of its harmonics. This property can also be exploited as a tool to select the artifact-related components from the ICA decomposition. To this end the heart rate frequency was determined by computing the inverse of the average time between consecutive QRS

complexes. Next it was checked whether the frequency spectrum of every component (computed with a normalized power spectral density (PSD) based on a Welch window) showed a peak at this frequency and its first harmonic. If these peaks were present, the component was identified as BCG-related.

c) Autocorrelation

Visually, the most apparent property of BCG-related independent components is their repetitiveness. A suitable measure for this repetitiveness is the autocorrelation function. The autocorrelation is the cross-correlation of a signal with a delayed version of itself and is often used for finding a repeating pattern in a signal (e.g., Deburchgraeve et al., 2008). BCG-related components can thus be distinguished by the peaks shown in their autocorrelation at exactly the distance between two consecutive QRS complexes.

d) Variance contribution

In Debener et al. (2008), another alternative selection procedure has been suggested, based on the amount of artifact variance explained by each of the independent components. Therefore, a matrix focusing on the artifactual activity (by taking epochs around the artifact occurrences) was used and the amount of variance explained in this matrix by every source was assessed with the EEGLAB function 'eeg_pvaf'. In the end all variances were divided by their sum in order to retrieve variance contributions in terms of percentages which are easier to threshold. Minimal variance contributions of both 5 and 10% were eventually selected as a criterion for the BCG identification.

e) Peak-to-peak of reconstructed sources

Another way to assess the characteristics of a specific source is to look at the signals obtained by backprojecting this source to the signal space. The criterion proposed here relies on the assumption that such backprojected signals should, if accounting for the BCG artifact, show patterns occurring at the same rhythm as the QRS complex in the ECG channel. Therefore, for each source, the reconstructed signals were epoched around the artifact occurrences and averaged. Of the resulting averages peak-to-peak (PTP) values were computed as the difference between the maximal and the minimal value of these averages in every channel. These peak-to-peak values were then divided by the peak-to-peak values of the original signals and the maximal value over all channels was taken. The obtained percentages gave an idea on which sources specifically cause high artifact amplitudes in the data. As thresholds 15 and 25% of the relative peak-to-peak amplitude were selected in this study.

OBS-ICA method

Although the majority of the combined EEG-fMRI studies rely solely on one of the two above procedures, Debener et al. (2005) proposed to combine them to optimize the artifact removal. More specifically, they suggested applying ICA to remove residual artifacts after BCG cleaning with OBS.

Because of the high quality of the results demonstrated with this combined approach, it was also included in our comparison study. Moreover, since the focus here was on an appropriate parameter selection, also the combination of all parameters needed to be investigated. These parameters are all the parameters explained in the separate OBS and

ICA sections. Slight modifications of certain threshold values were, however, necessary because the preceding OBS alters the data properties and were performed based on visual inspection of several examples. More specifically, for the correlation with the ECG channel, 0.07 and 0.15 were taken as thresholds and for the correlation with the mean and the mean squared BCG templates, 0.03 and 0.08 were chosen. Furthermore, we also wanted to account for subject-dependent differences in correlation, so a third threshold for the three correlation approaches was set on 75% of their maximal correlation. For the peak-to-peak value of the reconstructed sources 50 and 80% were chosen and for the variance contribution the thresholds became 5 and 15%.

Validation criteria

To assess the quality of the artifact removal and investigate the importance of the parameters, it is crucial to work with validation criteria that represent the different aims pursued with the data. Given the number of parameters, assessed validation was only feasible with quantified criteria. More specifically, for ICA alone, the number of evaluations per subject, task and block was 8 (possibilities for the number of components) x 4 (ICA algorithms) x 15 (selection criteria with different thresholds), equaling 480. For OBS-ICA this number needed to be multiplied by 7 (number of principal components), which resulted in 3320 possibilities.

The first objective of artifact removal is of course to remove the artifact as much as possible from the data. To this end the peak-to-peak value of the artifact-related epochs and the variation among single trials were studied. Of equal importance, however, is the requirement that the artifact removal should not affect the data of interest, in this case the

ERPs. Therefore, these ERPs were compared with the ones retrieved outside the scanner and their most important properties were studied. As opposed to previous similar studies, these properties were not only computed on averages over many trials, but also trial-to-trial characteristics were investigated.

1. Peak-to-peak value of artifact-related epochs

Our first validation criterion measures the amount of residual BCG artifact after the application of the different cleaning methods. A popular approach for this purpose is to look at certain properties of the averages of epochs retrieved around the QRS onset (e.g., Niazy et al., 2005; Mantini et al., 2007; Debener et al., 2008). Following Mantini et al. (2007), we extracted epochs from the EEG from 100 before until 900 ms after the QRS onset and subsequently averaged these. Then the peak-to-peak amplitudes of the averaged epochs were computed by subtracting the minimal from the maximal value in each channel, and averaged over all channels. Finally, a ratio was calculated of the resulting values before and after BCG cleaning. As such this criterion gives an idea on the reduction of the artifact compared to the original situation.

2. Single trial variation

A drawback of the above criterion lies in its assumption that the residual artifacts are not random and can thus be captured with an average measure. However, the residuals left by the algorithm can vary between occurrences (Niazy et al., 2005). Therefore, the use of a second criterion that pays more attention to this variability is mandatory. We defined a measure based on the variability among task-related single trials. The approach relied on

the assumption that, even though there can be small differences, task-related single trials should be largely similar. The presence of the artifact however, which is randomly spread over these trials and has varying amplitudes, causes the single trials to be more diverse. Therefore, the better the artifact is removed, the smaller the single trial variation will become. This criterion was implemented in the following way. Firstly, epochs were extracted per stimulus for both the detection and the Go/NoGo task. From all epochs the mean of the prestimulus interval was subtracted (baseline subtraction) and the epochs were slightly shifted in time to fit with their average. This fit was defined as the maximal cross-correlation, in order to account for small inaccuracies in time. Next the standard deviation over the trials was computed and averaged over all time points and channels. The eventual criterion is the ratio of the resulting values before and after BCG removal.

3. ERP difference

The first criterion to assess the effect of artifact removal on the data of interest compares the resulting ERPs with ERPs obtained with the same stimuli outside the scanner environment. To this end, Mantini et al. (2007) defined a difference measure as the square root of the mean squared difference between all time points of the averaged ERPs inside and outside the scanner. In our specific case, it was chosen to retrieve ERPs from both the Go and NoGo stimuli in the Go/NoGo task and from the central stimulus and all stimuli together in the detection task. To create the ERPs, firstly data were re-referenced to the mastoids and all blocks per subject and task were taken together. Then epochs were extracted and baseline subtraction was applied, after which trials with extreme values (above 200 μV) were removed and the average epoch computed. The difference measure

was computed following the formula in Mantini et al. (2007) and this difference was averaged over all channels.

4. Signal-to-noise ratio

Assecondi et al. (in revision) show that ERPs from inside and outside the scanner, are not necessarily similar, although the exact cause of this dissimilarity still needs to be investigated. Therefore, the ERP difference might not be completely representative for the quality of the ERPs. For this reason, Debener et al. (2007) proposed to use the signal-to-noise ratio (SNR) of peaks of interest in the averaged ERPs to assess the quality of the ERPs. They defined the SNR, following common clinical practice, as the amplitude of the selected peaks divided by the standard deviation of their prestimulus interval (200 ms before the stimulus onset). In the present study, the peaks used for this criterion were the P1 and N1 generated by the central stimulus in the detection task and the N2 and P3 generated by the Go and NoGo stimuli in the Go/NoGo task. The exact procedure to create the average ERPs was the same as for the ERP difference.

5. Task-related modulations

An important field of interest for ERPs is the effect of the stimulus type on properties of typical ERP components. The feasibility to use data for this purpose can therefore be considered as an additional criterion to evaluate the quality of the data. In the present study two examples of these "modulations" were studied. Firstly in the detection task, we chose the difference between responses on left and right visual field stimuli, without making a distinction between upper and lower parts. More specifically, the focus was on

the fact that the contralateral P1 (i.e. the P1 peak at the opposite side of the head as the stimulus) precedes the ipsilateral P1 (Di Russo et al., 2002). To obtain a measure for this, averaged ERPs were created for both left and right stimuli and an average from 5 occipital channels on each side of the head was retrieved. Subsequently the P1 was detected on the contralateral ERP and the amplitude of the ipsilateral ERP at the same latency was retrieved. A t-test was performed on the resulting contra- and ipsilateral amplitudes from all subjects to verify the significance of their difference. In the Results section the average amplitude difference over all subjects is reported (as an indication for the effect size) together with the p-value from the performed t-test. For the Go/NoGo task similar measures can be retrieved by looking at the N2 component on channel FCz. This N2 should be more pronounced in NoGo compared to Go conditions.

6. Trial-to-trial characteristics in the detection task

Traditionally, ERPs are extracted using an average over a large amount of trials. However, to fully exploit the advantages of combined EEG-fMRI measurements, it is more interesting to investigate the ERPs on a trial-to-trial basis. Since the above validation criteria are based on the average ERPs, they cannot detect small method-related differences in data quality that might affect these trial-to-trial dynamics. For this reason, there is a need for an additional measure that investigates the quality of the data on a much lower level of averaging.

Novitskiy et al. (in preparation) proposed to use single trial P1 and N1 amplitudes as regressors for fMRI analysis. More specifically, the amplitude differences in left and

right hemisphere electrodes related to the position of the stimulus on the visual field were employed.

A similar approach can be followed here to study the ERP quality in more detail. First of all, ERPs for both left and right stimuli were extracted following the same procedure as in the modulation criterion, meaning that upper and lower stimuli on each side of the visual field were taken together. The set of trials after baseline removal and extreme value rejection was used in two ways. First, the average over 5 left occipital channels for all right visual field stimuli was taken together with the average over 5 right occipital channels for the left visual field stimuli. As such, an average contralateral ERP was created allowing determining the latency of the contralateral P1. Secondly, the trial-to-trial quality of the set of trials was augmented by applying a Gaussian moving average (every window contained 11 trials). Afterwards, the amplitude of these trials at the P1 latency obtained in the first step was retrieved from the 5 left and 5 right occipital channels and averaged. As such, series of P1 amplitudes were obtained for both the left and the right hemisphere. Eventually, the right series were subtracted from the left. Since P1 amplitudes are known to be augmented on contralateral channels, the resulting values should be higher than zero for right visual field stimuli and lower than zero for the left ones.

The actual validation criterion measures for each stimulus type (left and right) the number of trials where the sign of the P1 difference value corresponds to the stimulus type, divided by the total number of trials. The resulting ratios represent two aspects of data quality. In the first place, the extreme value rejection removes all trials of insufficient quality by putting a threshold on the maximal amplitude of every trial.

Therefore the ratios will already be lower in cases with a lot of residual artifact. In the second place, both residual artifact and overcleaning of the data will deteriorate the relation between the P1 amplitude and the stimulus type and thus also lower the ratios. To establish the meaningfulness of these classification ratios, also an additional ratio was investigated for both left and right stimuli. These ratios were defined as the number of correctly classified trials divided by the real number of trials that is subjected to the classification, i.e., after trial rejection. These ratios should be significantly higher than 50%, since otherwise these numbers could also be retrieved by chance.

Statistical analyses

Due to the size of the present study, several statistical tests were needed. First of all, to study the main effects of all parameters in each of the methods, a repeated measurements analysis of variance (ANOVA) was performed, using the SAS 9.2 GLM procedure (SAS Institute Inc., Cary, NC, USA). More specifically, univariate tests were used of which the outcome was adjusted with the Huynh-Feldt Epsilon (H-F) correction factor, to account for possible violation of the sphericity assumption (Stevens, 1996).

In cases where only two sets of subject values had to be compared, paired two sample t-tests were used (after verification of normality and variance assumptions). More t-tests, but in a one sample version, were used to check whether the ratios of correctly classified trials divided through the number of good trials were significantly higher than 50%. These latter ratios were also evaluated on an individual basis through a proportion test introduced by Wilson et al. (1927).

Results

The results are presented in different sections. Since two of our validation criteria have not yet been used in related studies, their values are first shown for data acquired outside the MR scanner. Subsequently, simulation results are presented for the three employed approaches. The two main approaches (OBS and ICA) are then evaluated each separately on the experimental data with focus on parameter selection. The next results section is devoted to the comparison of these two methods using each method's most optimal parameter settings. In the last section, also the additive benefit of applying OBS prior to ICA is investigated on the experimental data, with special attention to subject-related differences.

Outside results

Two of the validation criteria employed in this study, namely task-related modulations and numbers of correctly classified trials were new to validate BCG removal methods and therefore it was necessary to assess their performance and illustrate their values on data without artifacts.

For the detection task data measured outside the MR scanner the average modulation across subjects was found to be 1.46 μV with a p-value of 0.0017. For the outside data of the Go/NoGo task this average modulation was 4.63 μV ($p = 0.0053$). These values show that task-related modulations should not only be present after good BCG cleaning, but should also be significant across subjects for both the detection and the Go/NoGo task.

As explained in the Methods section, the last criterion studies the probability of predicting the side of stimulus presentation, based on trial-to-trial P1 amplitude

asymmetry. For the left and right visual field stimuli these percentages, averaged over subjects, were respectively 74.34% and 73.75%. The p-values accounting for the fact that the pure classification part (i.e., the proportion of correctly classified trials in which only trials without extreme values are taken into account) should be better than chance (50%) were 0.0018 and 0.0181, respectively. These outside results demonstrate that data nicely cleaned from BCG artifacts should correspond to high classification percentages, not retrieved by only pure chance. They also provide us with an upper limit of the classification performance of realistic single trial data.

Simulation results

To form an initial idea of the differences in performance between OBS, ICA and OBS-ICA, it is interesting to have a look at the results of our simulation study. Figure 2 shows the root mean squared values of the differences between the original ERPs and the ERPs obtained after BCG cleaning with OBS, ICA and OBS-ICA for different noise levels (different SNR values on the horizontal axis). For BCG cleaning with OBS several numbers of principal components can be seen (ranging from 2 to 4) and for ICA and OBS-ICA small sets of parameter settings have been chosen as illustration. First of all, one can see that the presented curves show the expected behavior, in particular an increasing error for decreasing SNR values, up to some exceptions in the ICA graphs. For OBS, choosing different values for the number of principal components did not strongly affect the performance. Conversely, careful parameter selection appeared to be required for ICA since only with specific ICA parameters, errors could be found as low as with

OBS. With OBS-ICA these error values could even further be improved, on condition that proper parameter values were chosen.

The results from the validation on the realistic experimental data will now be discussed in the following sections.

OBS results

For pure OBS we were mainly interested in knowing whether the recommended number of principal components for the BCG removal (NPC = 3) was a good choice and whether OBS resulted in ERP data of sufficient quality. For this purpose, the main effect of the number of principal components, evaluated with the first four validation criteria, is summarized in Table 1.

The effect of the number of principal components on the peak-to-peak value is also shown for all subjects in Figure 3 for the detection task. For the Go/NoGo task the results looked similar.

The fact that for OBS the peak-to-peak value did not substantially decrease beyond 4 or 5 principal components indicates that in the remaining principal components artifactual activity is no longer present. This finding is not inconsistent with the fact that the single trial variation continued decreasing beyond this point, since also by removing non-artifactual activity the variation will decrease. The ERP difference was difficult to interpret here since the effect of the number of principal components was dependent on the stimulus under study. For the SNR no significant effect of the number of principal components could be found.

For the fifth evaluation criterion, i.e. the quality of task-related modulations in both tasks, the effect of the number of principal components is summarized in the upper part of Table 2. In the lower part of this table, the effects according to the sixth criterion, the trial-to-trial behavior of the data in the detection task, can be found.

Based on the first four criteria, which bear closest resemblances to earlier studies, 2, 3 or 4 principal components would give the best results. However, according to the task-related modulation criterion it can be concluded that only NPC = 3 and NPC = 4 yield a difference consistent over all subjects for the detection task. For the Go/NoGo task all selected numbers of principal components give significant differences but the best value is obtained with NPC = 2. For the trial-to-trial percentages only with NPC = 2 and NPC = 3 values are found of which the pure classification part is significantly higher than 50% for both left and right visual field stimuli. These values are however Thus, taken together the findings from all six criteria, only OBS with three principal components gives good and significant values for each of the criteria.

ICA results

First of all it was tested whether performing ICA on selected (artifact-related) parts of the data instead of on the whole dataset would improve or speed up the artifact removal. However, no significant effect with any of the proposed validation criteria could be found. Moreover, since BCG artifacts were prominent during nearly the whole time course of the data, this data selection did not give a significant time reduction either. Therefore all presented results are based on ICA performed on the continuous EEG data.

To be able to extract the most interesting findings from the data, the evaluation was performed in two stages. First, a subset of good parameter settings was generated based on the first four validation criteria and on visual inspection of grand average ERPs. Second, this subset was evaluated with the task-related modulation and the number of correctly classified trials.

The three parameters of interest in ICA were: algorithm, number of components and selection criterion. All three significantly affected each of the first four validation criteria in both tasks. To summarize these results in an insightful manner, thresholds were set on each criterion, distinguishing acceptable from really poor results. The thresholds were based on the values retrieved with the same criteria after OBS. More specifically, the lowest values retrieved with OBS were taken (as a kind of a worst case OBS), except for the peak-to-peak value for which a less stringent threshold was chosen. Afterwards, all parameter combinations that yielded values above the thresholds for all four of the validation criteria were selected. The same parameter combinations were also evaluated with visual inspection of the grand average ERPs. Neither the values themselves nor the most optimal value are shown here quantitatively since discussing them in detail would be too elaborate for the scope of this paper. Moreover optimal values seemed to differ between different criteria. Therefore, in Table 3 only the parameter combinations retained after these two steps (thresholding and visual inspection), are presented.

Of the remaining parameter combinations, Figure 4 shows the amplitudes of task condition induced ERP modulations together with their significance, for JADE (Figure 4a), SOBI (Figure 4b), FastICA (Figure 4c) and Infomax (Figure 4d) respectively. Also the number of correctly classified trials was evaluated and these cases where the

proportion of correctly classified trials (relative to the number of trials after rejection of extreme values) was significantly higher than 50% are presented in Table 4.

When the first four validation criteria are considered (Table 3), FastICA seems to require a higher number of components, while an optimal choice in selection criterion is less essential. This effect was confirmed by visual inspection, revealing that choosing a number of components lower than 50, completely deteriorates the results. For JADE and SOBI the effect is rather the opposite: all numbers of components yield similar results as long as a proper selection criterion is used. Infomax seems to be less stringent in the choice of numbers of components and selection criteria. Only very low numbers of components (15 and 20) and two of the selection criteria (correlation with mean squared BCG channel and frequency content) give bad results. In all other cases, acceptable ERP signals can be retrieved, although in many cases the amplitude of the Go and NoGo peaks seems to be reduced (this is not shown here).

When looking at the task-related modulations (Figure 4), also JADE, SOBI and Infomax can only obtain better modulation results when using the highest number of components, thereby loosing their advantage over FastICA. For all four ICA algorithms the Go/NoGo N2 modulation is significant in many cases, regardless of a good parameter choice, which is in line with the OBS results. The P1 modulation in the detection task, however, is significant in only a few cases with JADE, SOBI and Infomax. This is probably because the P1 peak is much smaller than Go/NoGo N2 and thus more sensitive to residual artifact and overcleaning. Infomax seems to better preserve the P1 modulation, but the amplitude differences of the N2 are reduced, which is in correspondence with the abovementioned reduction of Go and NoGo peak amplitudes.

The sensitivity of the P1 peak is even further confirmed with the last criterion. Only five parameter combinations allowed significantly above chance classification of side of stimulus presentation, based on trial-to-trial P1 amplitude asymmetry (Table 4).

Comparison of OBS with ICA

An example of grand average ERPs obtained after OBS with 3 principal components is given above in Figure 5 (panel a: Go/NoGo, panel b: Detection task). The corresponding grand averages for ICA are shown in the centre of Figure 5 (panels c and d). The ICA parameters (FastICA, 62 components and peak-to-peak value of reconstructed sources as selection criterion), were not the most optimal, but merely serve as an illustration, as one of the good parameter choices (Table 3). As a reference also the grand average ERPs obtained outside the scanner are shown below in Figure 5 (panels e and f).

From these figures it can be seen that both with OBS and ICA it is possible to obtain high quality grand average ERPs that show peaks and task-related modulations known from previous related studies. Moreover, the obtained components correspond to findings on data measured outside the scanner.

Table 5 shows for each of the first four validation criteria a quantitative comparison between OBS and ICA with, for each ICA algorithm, parameters selected from the subset in Table 3. In particular, parameter sets were chosen for which the values of the ERP difference and the SNR of the ICA results are not significantly different from those of the OBS results. Apart from that, the presented parameter combinations are neither the most optimal nor the only good choices, but serve as an illustrative comparison of OBS and ICA. Table 5 shows that comparable results can by obtained with OBS and ICA for all

validation criteria. Moreover, comparing Table 2 with Table 4 and with Figures 4a, 4b, 4c and 4d also reveals that similar values for task-related modulations and number of correctly classified trials can be found with OBS and ICA. It should be noted, however, that to obtain this quality of results with ICA, it is important to carefully optimize the parameter settings.

OBS-ICA results

The additional application of ICA after OBS implied a vast amount of parameters to set and was therefore clearly more complicated than pure OBS. For this reason the focus in this part was particularly on those OBS-ICA results outperforming those of OBS alone. The peak-to-peak value and single trial variation are not very useful for this purpose since we know that with OBS alone already good values can be obtained. Moreover, the danger of the additional ICA is removing too much task-related signal from the data and this cannot be detected with these validation criteria. From our experience, also the ERP difference is not specific enough to pick up the differences and therefore it was decided to select a subset of suitable OBS-ICA parameters based on the SNR and the modulation values. To this end, all parameter combinations were retrieved for which the SNR was not significantly worse than that of OBS and that yielded a significant task-related modulation for both the detection and the Go/NoGo task. The SNR may not be significantly worse since any additional ICA step should not decrease the amplitude of the peaks of interest. As such the subset shown in Table 6 was retrieved.

Of this subset the performance of classifying trials according to their P1 behavior was evaluated and all cases significantly better than in the pure OBS case are listed in Table

7. Significantly better is defined here as the fact that the minimum of the percentages of both left and right stimuli should be significantly higher than the corresponding percentage from OBS alone (e.g., if the percentage for the left stimulus is the lowest, then left percentages are compared) and the other percentage should not be significantly lower.

This table shows that it is indeed possible to obtain better trial-by-trial data quality by combining OBS and ICA but only with specific parameter settings. However, finding these parameter settings is not straightforward. Table 6 and 7 can give a general idea on which parameters to choose, but caution is recommended because, especially in Table 7, these parameters might be tuned specifically to our data. However, in our results JADE, SOBI and Infomax were found to be effective, especially with correlation-based selection criteria.

It should be noted that the effect of an additional ICA step can largely depend on the individual under study. This is illustrated in Figure 6 for the trial-by-trial classification percentages. For each of the subjects and five different OBS-ICA parameter combinations, the differences between the OBS and OBS-ICA percentages are plotted for both the left and right visual field stimuli. However, five different parameter settings were chosen, each of them increasing the percentages for one subject, while decreasing them for most other subjects. A positive value here means that the percentages are higher with the additional ICA step, while a negative value indicates that the additional ICA step decreases the quality of the results. As such, this figure shows that choosing appropriate parameter settings is essential for the performance of ICA after OBS and that ‘wrong’

choices can completely deteriorate the results. The individual analyses were also performed for different numbers of principal components for pure OBS. However, in that case the results were more consistent over subjects and similar trends as with OBS-ICA could not be found.

Discussion

In this study we compared several existing methods and their parameter settings for BCG artifact removal from EEG data recorded inside an MR scanner. More specifically, OBS, ICA and OBS-ICA were used to clean ERP data from a visual detection and a Go/NoGo task. Validation of the BCG removal was performed with six different criteria, not only characterizing the residual artifact but, more importantly, also looking at the quality of the ERPs. We found that, although parameter settings strongly affected the performance of each method, no one method emerged as the best current approach. In general, OBS and ICA can yield comparable results, but the quality of the ICA results strongly depends on the choice of parameters. The need for careful parameter tuning makes ICA therefore less interesting for non-experienced users and also for experienced users it is possibly very time-consuming. Combining OBS and ICA can increase data quality. However, we showed that, if wrong parameter choices are made, the additional ICA step might as well spoil the quality obtained with pure OBS.

The OBS approach, the most widely used subtraction method, uses PCA to create a template for each artifact occurrence in each channel. Our results show that the recommended value of three for the number of principal components is indeed most

optimal. Subtracting more principal components will not remove much artifactual activity and less than three components do not capture the considerable temporal variation of the BCG artifact. Moreover, only NPC = 3 gives significant values for more detailed validation criteria (task-related modulation and trial-to-trial behavior), although the classification performance is still less than in the case of outside data.

The application of ICA for the removal of the BCG artifact, attempts to separate both artifact- and brain-related sources by decomposing the EEG into independent components. Decomposition can be performed into several numbers of components with different algorithms, and various selection criteria can be used to distinguish the artifactual from the other sources.

In their study on the comparison of OBS and ICA, Debener et al. (2007) concluded that with OBS a better SNR could be obtained than with ICA. However, we found that by choosing the appropriate parameter settings, ICA can yield SNR values that are not significantly different from those of OBS. This is nevertheless not inconsistent with Debener's findings, since they already indicated themselves that the number of removed components was not optimized in their study.

Trial-to-trial behavior and task-related modulation criteria further showed that identifying appropriate ICA parameter settings is not trivial. It is however important, because with many parameter combinations it would not be possible to retrieve the expected ERP waveforms. Although the results are not always straightforward, our study reveals important overall patterns. Reducing the dimensionality of the data before applying ICA appeared in many cases not to be beneficial and even harmful for the data quality,

especially when using FastICA. With the other algorithms this effect was less prominent, although a solid explanation for this difference could not be found. Considering the selection of components, some selection criteria were preferable to others. Correlation-based measures proved to be effective for selecting the major artifact components, but also the variance contribution, as suggested by Debener et al. (2008), and the peak-to-peak value of reconstructed sources are reliable selection criteria for ICA. Moreover, these three criteria have the advantage of yielding acceptable average results regardless of the chosen ICA algorithm. The nice performance of the variance contribution and the peak-to-peak value of reconstructed sources can probably be explained by the fact that BCG artifacts can be very effectively identified by looking at properties of the original data. The obvious rhythm of the BCG artifact (at the same rate as the ECG signal) is clearly not so pronounced in the artifactual independent components, reducing the performance of the frequency content and autocorrelation criteria.

When comparing OBS and ICA, it should therefore be clear that the extensive parameter tuning makes ICA less attractive for BCG removal than OBS. With OBS, acceptable results are more easily obtained with the default number of principal components. OBS seems also more robust to different noise levels (see simulation study) than ICA. Furthermore, despite of all the effort and time spent in finding optimal ICA parameter settings, results were never significantly better than those of OBS.

We also investigated the combination of OBS and ICA. Confirming the findings of Debener et al. (2007), the additional use of ICA after OBS can indeed increase the SNR.

In addition, also for our other validation criteria OBS-ICA can yield an improvement compared to pure OBS. However, this improvement can easily turn into deterioration, as was illustrated with the individual results. Thus, when applying ICA after OBS, one should always carefully check the quality of the resulting ERPs. Automating this additional step is, therefore, not very obvious.

Tracing back the individual differences in the additional ICA step to the characteristics of the BCG artifact in the data might have been useful in finding good parameter settings. As already indicated, the BCG artifact can largely depend on the subject and channel under study, as can also be retrieved with e.g. global field power values (Skrandies, 1990; Assecondi et al., 2009). However, strong relations between such BCG characteristics and the performance of the methods could not be found.

Besides the actual comparison of methods, this study also revises and introduces validation criteria for the evaluation of new methods. Although the perfect removal of the BCG artifact is still not within reach, the used criteria show that acceptable EEG quality can be obtained with the existing methods. As indicated above, OBS combines good data quality with user-friendliness and is therefore preferred to ICA, which requires much more parameter tuning. A possible improvement can be obtained with the combination of OBS and ICA, but, in any case, our results show that it is essential to check in sufficient detail the obtained data quality.

A first important quality check is certainly the averaged ERP. To this end, Mantini et al. (2007) proposed to use the difference between inside and outside ERPs. However, this ERP difference seems a rather rough measure for the quality of the ERPs, possibly due to

the fact that components from inside and outside ERPs appear to have different latencies (as also shown by Assecondi et al., in revision). Therefore, the ERP difference might indicate only large differences in ERPs. Moreover, computing the ERP difference also requires EEG data from the same subjects measured outside the scanner environment, thereby increasing the recording time. The ERP difference was, however, included in this study in order to be able to compare our results with other, similar studies in the literature (e.g., Mantini et al., 2007) and evaluate the adequacy of this previously suggested validation criterion. The SNR (as proposed by Debener et al., 2007) is a better validation criterion, although residual artifacts could possibly wrongly be regarded as peaks of interest when automated peak detection is used. For this reason, studying the SNR should always be combined with visual inspection of the averaged ERPs.

Previous validation studies, using validation criteria based on averaged ERPs, were not able to reveal the benefit of using three principal components for OBS, as shown here. Furthermore, the importance of extensive parameter tuning in ICA would also not have been revealed with these average criteria. Therefore, in order to make a distinction between the methods under study clearer than what was previously possible, we included very detailed quality checks in our validation study.

Although many ERP studies investigate properties of ERP components related to different conditions during the experimental task, preserving these task-related modulations was, to our knowledge, never employed before to investigate the quality of the BCG removal. Computing these same criteria on data acquired outside the magnetic field of the scanner however proved their ability to reveal good ERP data quality and indicated upper limits for what can be expected from these measures.

We believe that in most ERP studies, it should in general be possible to evaluate some kind of these task-related modulations on a trial-by-trial level, as was done in this study on the detection task data. Since single trial data already suffer from a low SNR outside the MR environment, additional denoising (with e.g. the Gaussian average used in this study) might nevertheless be necessary. As shown in this paper, the use of task-specific and trial-to-trial characteristics can therefore provide the user with a more widely validated quality assurance.

With our improved validation criteria, our findings extend the results of previous studies. For instance, Grouiller et al. (2007) directly compared OBS and ICA, but their validation was performed on simulation data, alpha activity and epileptic spikes. Although the characteristics of these EEG patterns are completely different from those of ERPs, OBS proved to be a good choice in their study as well. This effect was strongest when the BCG artifact strongly contaminated the recordings, as is the case at the high field strength of 3T. Moreover, this is in agreement with the conclusions of our simulation study, namely that also high artifact amplitudes (low SNRs) still lead to reasonable error values with OBS.

The comparison of BCG removal methods in this study focuses on OBS- and ICA-based methods as these are most widely used. One of the first BCG removal methods, the Average Artifact Subtraction (AAS) might therefore seem not to be included in our comparison. However, since the first principal component obtained with OBS is always the average across artifact occurrences, using OBS with only 1 principal component can

be seen as a type of AAS. It should nevertheless be noted that this implementation of AAS is not the most optimal one, since true AAS uses sliding windows, therefore partially accounting for the variability of the artifact.

As already indicated, also other methods emerged recently, in particular the CCA-based BCG removal method as described in Assecondi et al. (2009). Since the latter method was not well established and validated yet on ERPs, we decided not to include this method in the current study. However, we refer the interested reader to a separate paper (Assecondi et al., in revision), in which the BCG removal effects of this novel method are more fully investigated on ERPs including the same datasets as in the current study. In addition, Assecondi et al. (in revision) also studies the effect of the magnetic field on the characteristics of the ERPs.

Conclusion

Removing the BCG artifact from EEG data remains a major bottleneck for a successful integration of simultaneously recorded EEG and fMRI data, because the artifact shows a considerable variation over time and channels. Many attempts have been made, of which the most widely used ones, OBS and ICA as well as a combination of both, were compared in this study. Special care was taken to evaluate a large choice of possible parameter settings, combined with a comprehensive selection of informative validation criteria including task-related measures in single trial analysis. While the traditional validation criteria, based on average characteristics, were not specific enough to point out the difference between both methods, our task-specific criteria provided us with a much clearer distinction between OBS and ICA. The advantage of looking at stimulus-related

ERP properties both on an average as on a trial-to-trial basis, is twofold. Firstly, we go beyond the typical averaging in ERPs, which improves data quality and smoothes specific differences between methods. Secondly, these task-related modulations and the use of ERP data on a single trial basis are major research topics of interest for ERP scientists. Both OBS and ICA proved to yield equally good results, but since these results are much harder to obtain with ICA, OBS is a better choice for a good BCG artifact cleaning, in particular for beginners and less experienced users in the field. Indeed, OBS is less prone to inter-subject variations and more easy to use, thereby enabling automation and a more general use in clinical practice. However, since OBS also leaves some residual artifact, one might also consider using an additional ICA step after OBS. In this case it is essential to carefully monitor the removed activity and this for all subjects under study, because results might be very different between individuals. Applying ICA after OBS might as well improve as completely deteriorate the results.

In conclusion, these results provide important insights into how to optimize settings for OBS, ICA and OBS-ICA. These findings should yield enough informative guidelines for experts as well as newbies in the field to obtain the desired data quality in simultaneous EEG and fMRI measurements.

Acknowledgements

The authors would like to thank Ing. Sven Gijsen, MR technician at the Maastricht Brain Imaging Centre for his kind help with the equipment and experiments. They are also grateful to Dr. Maarten Mennes for his useful advice on different stages of the data analysis.

This research is supported by the Research Council KUL: GOA-AMBioRICS, GOA MaNet and IDO 05/010 EEG-fMRI; the Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, 'Dynamical systems, control and optimization', 2007-2011) and the EU project Neuromath (COST-BM0601). Katrien Vanderperren is supported by a PhD grant from the Flemish Government (IWT). Nikolay Novitskiy is supported by a post-doctoral scholarship of Research Foundation – Flanders (FWO).

References

Allen, P.J., Polizzi, G., Krakow, K., Fish, D.R., Lemieux, L., 1998. Identification of EEG events in the MR scanner: the problem of pulse artifact and a method for its subtraction. NeuroImage 8, 229–239.

Allen, P.J., Josephs, O., Turner, R., 2000. A method for removing imaging artifact from continuous EEG recorded during functional MRI. NeuroImage 12, 230–239.

Assecondi, S., Hallez, H., Staelens, S., Bianchi, A., Huiskamp, G., Lemahieu, I., 2009. Removal of the ballistocardiographic artifact from EEG-fMRI data: a canonical correlation approach. Physics in Medicine and Biology 54 (2), 1673–1689.

Assecondi, S., Vanderperren, K., Novitskiy, N., Ramautar, J., Fias, W., Stiers, P., Sunaert, S., Staelens, S., Van Huffel, S., Lemahieu, I., in revision. Effect of the static magnetic field of the MR-scanner on ERPs: evaluation of visual, cognitive and motor potentials.

Babiloni, F., Mattia, D., Babiloni, C., Astolfi, L., Salinari, S., Basilisco, A., Rossini, P.M., Marciani, M.G., Cincotti, F., 2004. Multimodal integration of EEG, MEG and fMRI data for the solution of the neuroimage puzzle. Magn. Reson. Imaging 22 (10), 1471–1476.

Bell, A. J., Sejnowski, T. J., 1995. An information-maximization approach to blind separation and blind deconvolution. Neural Computation 7(6), 1129–59.

Belouchrani, A., Abed-Meraim, K., Cardoso, J.F., Moulines, E., 1993. Second-order blind separation of temporally correlated sources. Proc. Int. Conf. on Digital Sig. Proc., (Cyprus), 346–351.

Bénar, C.G., Aghakhani, Y., Wang, Y., Izenberg, A., Al-Asmi, A., Dubeau, F., Gotman, J., 2003. Quality of EEG in simultaneous EEG–fMRI for epilepsy. Clin. Neurophysiol. 114, 569–580.

Bokura, H., Yamaguchi, S., Kobayashi, S., 2001. Electrophysiological correlates for response inhibition in a Go/NoGo task. Clin. Neurophysiol. 112 (12), 2224–2232.

Cardoso, J.F., Souloumiac, A., 1993. Blind beamforming for non-Gaussian signals. IEE Proc.-F. 140 (6), 362–370.

Comon, P., 1994. Independent component analysis—A new concept? Signal Process. 36, 287–314.

Czisch, M., Wetter, T.C., Kaufmann, C., Pollmacher, T., Holsboer, F., Auer, D.P., 2002. Altered processing of acoustic stimuli during sleep: reduced auditory activation and visual deactivation detected by a combined fMRI/EEG study. NeuroImage 16, 251–258.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D.Y., Engel, A.K., 2005. Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. J. Neurosci. 25 (50), 11730–11737.

Debener, S., Ullsperger, M., Siegel, M., Engel, A.K., 2006. Single-trial EEG–fMRI reveals the dynamics of cognitive function. Trends Cogn. Sci. 10 (12), 558–563.

Debener, S., Strobel, A., Sorger, B., Peters, J., Kranczioch, C., Engel, A.K., Goebel, R., 2007. Improved quality of auditory event-related potentials recorded simultaneously with 3T fMRI: Removal of the ballistocardiogram artifact. NeuroImage 34, 587–597.

Debener, S., Mullinger, K.J., Niazy, R.K., Bowtell, R.W., 2008. Properties of the ballistocardiogram artefact as revealed by EEG recordings at 1.5, 3 and 7 Tesla static magnetic field strength. Int. J. Psychophysiol. 67 (3), 189–199.

Deburchgraeve, W., Cherian, P.J., De Vos, M., Swarte, R.M., Blok, J.H., Visser, G.H., Govaert, P., Van Huffel, S., 2008. Automated neonatal seizure detection mimicking a human observer reading EEG. Clin. Neurophysiol. 119 (11), 2447–2454.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Meth. 134, 9–21.

Di Russo, F., Martinez, A., Sereno, M.I., Pitzalis, S., Hillyard, S.A., 2002. Cortical sources of the early components of the visual evoked potential. Hum. Brain Mapp. 15, 95–111.

Eichele, T., Calhoun, V.D., Moosmann, M., Specht, K., Jongsmae, M.L.A., Quian Quiroga, R., Nordby, H., Hugdahl, K., 2008. Unmixing concurrent EEG-fMRI with parallel independent component analysis. Int. J. Psychophysiol. 67(3), 222–234.

Ellingson, M.L., Liebenthal, E., Spanaki, M.V., Prieto, T., Binder, J.R., Ropella, K.M., 2004. Ballistocardiogram artifact reduction in the simultaneous acquisition of auditory ERPS and fMRI. NeuroImage 22, 1534–1542.

Falkenstein, M., Hoormann, J., Hohnsbein, J., 1999. ERP components in Go/Nogo tasks and their relation to inhibition. Acta Psychol. 101, 267–291.

Goldman, R.I., Stern, J.M., Engel, J.Jr., Cohen, M.S., 2000. Acquiring simultaneous EEG and functional MRI. J. Clin. Neurophysiol. 111, 1974–1980.

Goldman, R.I., Stern, J.M., Engel, J.Jr., Cohen, M.S., 2002. Simultaneous EEG and FMRI of the alpha rhythm. NeuroReport 13, 2487–2492.

Grouiller, F., Vercueil, L., Krainik, A., Segebarth, C., Kahane, P., David, O., 2007. A comparative study of different artifact removal algorithms for EEG signals acquired during functional MRI. NeuroImage 38, 124–137.

Huiskamp, G.J., 2005. Reduction of the Ballistocardiogram Artifact in Simultaneous EEG-fMRI using ICA. Conf. Proc. IEEE Eng. Med. Biol. Soc. 4, 3691–3694.

Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent component analysis. John Wiley & Sons.

Joyce, C.A., Gorodnitsky, I.F., Kutas, M., 2004. Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. Psychophysiology 41 (2), 313–325.

Kim, K.H., Yoon, H.W., Park, H.W., 2004. Improved ballistocardiac artifact removal from the electroencephalogram recorded in fMRI. J. Neurosci. Meth. 135 (1–2), 193–203.

Klemm, M., Haueisen, J., Ivanova, G., 2009. Independent component analysis: comparison of algorithms for the investigation of surface electrical brain activity. Med. Biol. Eng. Comput. 47, 413–423.

Krakow, K., Messina, D., Lemieux, L., Duncan, J.S., Fish, D.R., 2001. Functional MRI activation of individual interictal epileptiform spikes. NeuroImage 13, 502–505.

Laufs, H., Kleinschmidt, A., Beyerle, A., Eger, E., Salek-Haddadi, A., Preibisch, C., Krakow, K., 2003. EEG-correlated fMRI of human alpha activity. NeuroImage 19, 1463–1476.

Lee, T.W., Girolami, M., Sejnowski, T.J., 1999. Independent Component Analysis using an Extended Infomax Algorithm for Mixed Sub-Gaussian and Super-Gaussian Sources. Neural Computation 11(2), 417–441.

Lemieux, L., Krakow, K., Fish, D.R., 2001. Comparison of spike-triggered functional MRI BOLD activation and EEG dipole model localization. NeuroImage 14, 1097–1104.

Liebenthal, E., Ellingson, M.L., Spanaki, M.V., Prieto, T.E., Ropella, K.M., Binder, J.R., 2003. Simultaneous ERP and fMRI of the auditory cortex in a passive oddball paradigm. NeuroImage 19, 1395–1404.

Makeig, S., Debener, S., Onton, J., Delorme, A., 2004. Mining event-related brain dynamics. Trends Cogn. Sci. 8 (5), 204–210.

Mantini, D., Perrucci, M.G., Cugini, S., Ferretti, A., Romani, G.L., Del Gratta, C., 2007. Complete artifact removal for EEG recorded during continuous fMRI using independent component analysis. NeuroImage 34, 598–607.

Moosmann, M., Ritter, P., Krastel, I., Brink, A., Thees, S., Blankenburg, F., Taskin, B., Obrig, H., Villringer, A., 2003. Correlates of alpha rhythm in functional magnetic resonance imaging and near infrared spectroscopy. NeuroImage 20, 145–158.

Moosmann, M., Eichele, T., Nordby, H., Hugdahl, K., Calhoun, V.D., 2008. Joint independent component analysis for simultaneous EEG-fMRI: Principle and simulation. Int. J. Psychophysiol. 67 (3), 212–221.

Mulert, C., Jäger, L., Schmitt, R., Bussfeld, P., Pogarell, O., Möller, H.J., Juckel, G., Hegerla, U., 2004. Integration of fMRI and simultaneous EEG: towards a comprehensive understanding of localization and time-course of brain activity in target detection. NeuroImage 22, 83–94.

Niazy, R.K., Beckmann, C.F., Iannetti, G.D., Brady, J.M., Smith, S.M., 2005. Removal of FMRI environment artifacts from EEG data using optimal basis sets. NeuroImage 28, 720–737.

Nieuwenhuis, S., Yeung, N., Van den Wildenberg, W., Ridderinkhof, K.R., 2003. Electrophysiological correlates of anterior cingulate function in a Go/NoGo task: Effects of response conflict and trial-type frequency. Cog., Affect. Behav. Ne. 3, 17–26.

Novitskiy, N., Ramautar, J., Vanderperren, K., Vanrumste, B., Stiers, P., Van den Bergh, B., Lagae, L., Sunaert, S., Van Huffel, S., Wagemans, J., in preparation. The BOLD correlates of the visual P1 and N1 in single-trial analysis of simultaneous EEG-fMRI recording during a spatial detection task.

Schabus, M., Dang-Vu, T.T., Albouy, G., Balteau, E., Boly, M., Carrier, J., Darsaud, A., Degueldre, C., Desseilles, M., Gais, S., Phillips, C., Rauchs, G., Schnakers, C., Sterpenich, V., Vandewalle, G., Luxen, A., Maquet, P., 2007. Hemodynamic cerebral correlates of sleep spindles during human non-rapid eye movement sleep. P. Natl. Acad. Sci. USA 104, 13164–13169.

Seeck, M., Michel, C.M., Spinelli, L., Lazeyras, F., 2001. EEG mapping and functional MRI in presurgical epilepsy evaluation. Rev. Neurol. (Paris) 157 (8–9 Pt. 1), 747–751.

Sijbers, J., Van Audekerke, J., Verhoye, M., Van der Linden, A., Van Dyck, D., 2000. Reduction of ECG and gradient related artifacts in simultaneously recorded human EEG/MRI data. Magn. Reson. Imaging 18 (7), 881–886.

Skrandies, W., 1990. Global field power and topographic similarity. Brain Topogr. 3 (1), 137–141.

Srivastava, G., Crottaz-Herbette, S., Lau, K.M., Glover, G.H., Menon, V., 2005. ICA-based procedures for removing ballistocardiogram artifacts from EEG data acquired in the MRI scanner. NeuroImage 24, 50–60.

Stevens, J.P., 1996. Applied Multivariate Statistics for the Social Sciences, Third Edition. Mahway NJ, Lawrence Erlbaum Associates, Inc.

Stone, J.V., 2002. Independent component analysis: an introduction. Trends Cogn. Sci. 6, 59–64.

Tang, A.C., Sutherland, M.T., McKinney, C.J., 2005. Validation of SOBI components from high-density EEG. NeuroImage 25 (2), 539–553.

Urrestarazu, E., Iriarte, J., Alegre, M., Valencia, M., Viteri, C., Artieda, J., 2004. Independent component analysis removing artifacts in ictal recordings. Epilepsia 45 (9) 1071–1078.

Vanderperren K., Ramautar, J., Novitski, N., De Vos, M., Mennes, M., Vanrumste, B., Stiers, P., Van den Bergh, B., Wagemans, J., Lagae, L., Sunaert, S., Van Huffel, S., 2007. Ballistocardiogram artifacts in simultaneous EEG-fMRI acquisitions. International

Journal of Bioelectromagnetism, Special Issue on Methods for the estimation of brain activity I 9 (3), 146–150.

Wilson, E.B., 1927. Probable inference, the law of succession, and statistical inference. J. Am. Stat. Assoc. 22 (158), 209–212.

Yan, W.X., Mullinger, K.J., Geirsdottir, G.B., Bowtell, R., 2009. Physical Modeling of the Pulse Artefact Sources in Simultaneous EEG/fMRI. Hum. Brain Mapp. Published Online: 12 Oct 2009. DOI 10.1002/hbm.20891.

Figures

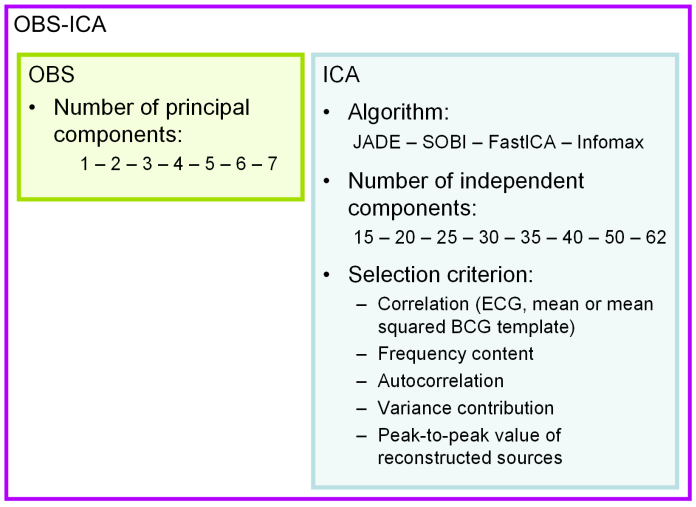


Figure 1: Overview of methods evaluated in this study for the removal of the BCG artifact together with the parameters investigated for each method. For the combined approach OBS-ICA, all parameters from both OBS and ICA were included.

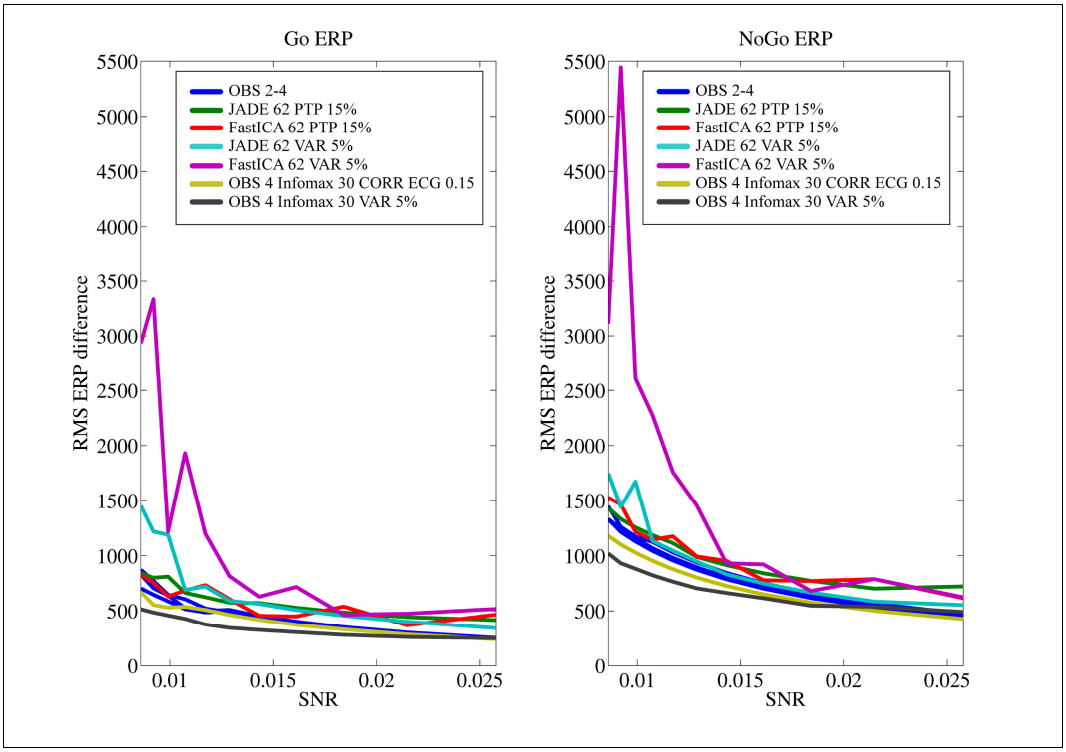


Figure 2: Root mean squared value of the difference between the original ERPs and the averaged ERPs obtained after BCG cleaning on simulated data with different noise levels. Values are shown for both the Go and the NoGo ERPs and for the OBS method (with number of principal components ranging from 2 to 4), the ICA approach and the combined approach OBS-ICA. For ICA the algorithms JADE and FastICA are compared with 62 components and the variance contribution (VAR) and the peak-to-peak value of reconstructed sources (PTP) as selection criteria. For OBS-ICA curves are shown for Infomax with 30 components performed after OBS with 4 principal components and with the correlation with the ECG (CORR ECG) and the variance contribution (VAR) as selection criteria.

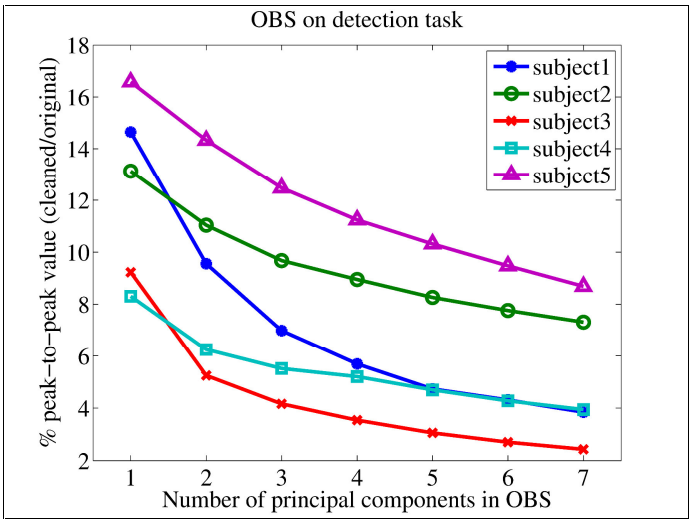


Figure 3: Ratio (in %) of peak-to-peak values of BCG related epochs after and before BCG cleaning with OBS. The results are shown for the detection task for the five subjects (in different colors) and for a number of principal components ranging from 1 to 7.

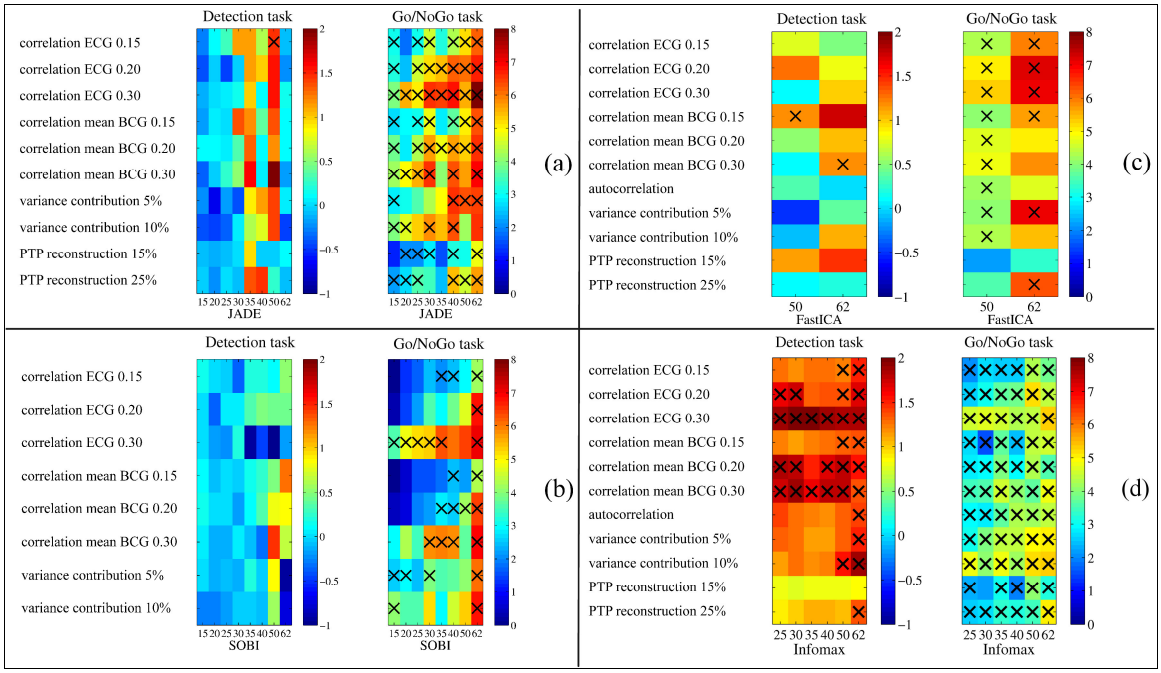


Figure 4: Modulation values illustrated with color scale with X sign added if the shown modulation is significant over subjects. The values are shown for the selection criteria

and number of components from the subset presented in table 4 for (a) JADE; (b) SOBI; (c) FastICA; (d) Infomax and for both the detection task (left) and the Go/NoGo task (right).

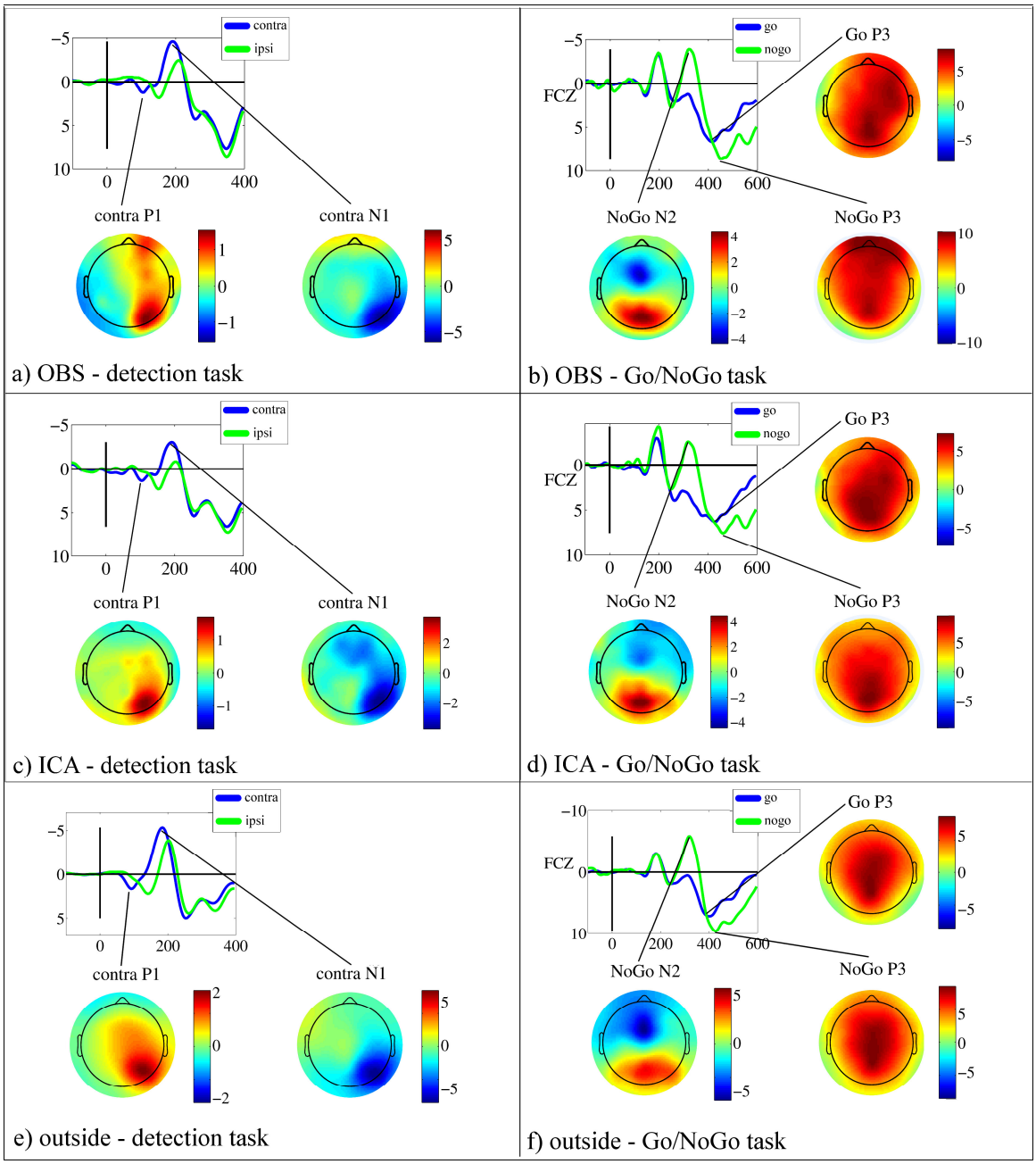


Figure 5: Grand average ERPs (from five subjects) after OBS with number of principal components equaling 3 (upper figures (a) and (b)) and after ICA (FastICA, 62 components, peak-to-peak value of reconstructed sources 25%) (centre figures (c) and (d)) with scalp maps of specific ERP components. Figures (e) and (f) show the same grand average ERPs and scalp maps obtained from the data measured outside the scanner. Figures (a), (c) and (e) Contra- (blue) and ipsilateral (green) ERPs for detection task in which 5 occipital channels on the left and right hemisphere were pooled together. Scalp maps are shown for the contralateral P1 and N1; (b), (d) and (f) Go (blue) and NoGo (green) ERPs for Go/NoGo task on frontocentral channel FCZ with scalp maps of the NoGo N2 and P3 and the Go P3.

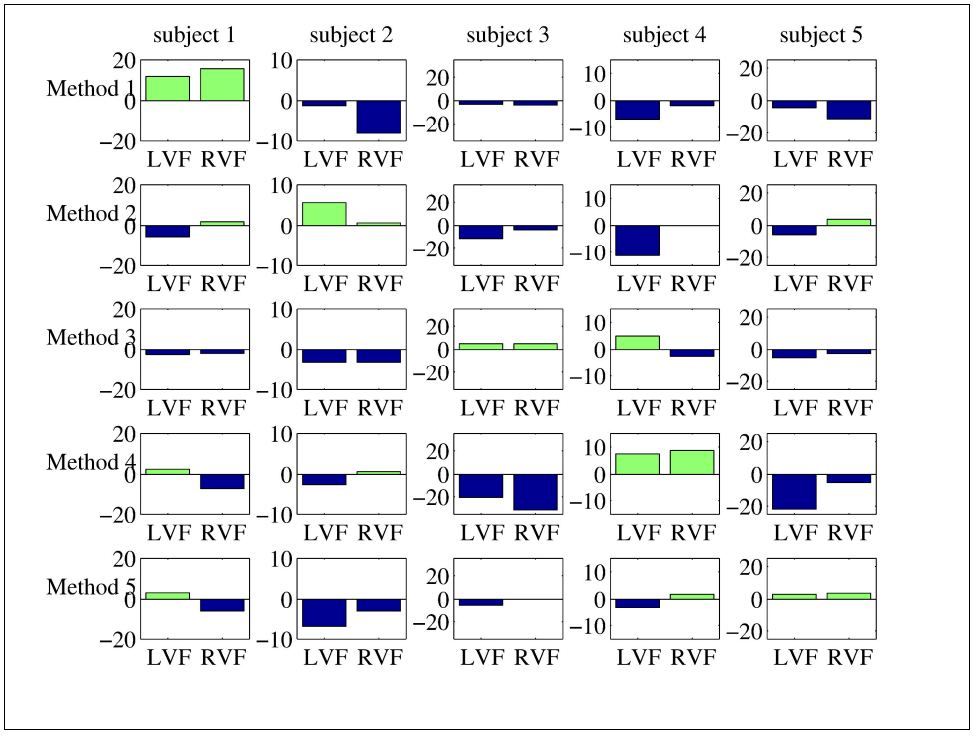


Figure 6: Difference between classification percentages of OBS-ICA with five different OBS-ICA parameter combinations and OBS with 3 principal components. Each column

represents a different subject and each row a different OBS-ICA parameter combination (=method). The percentage differences are shown for both left and right visual field stimuli (LVF and RVF). Negative values are shown in blue, positive values in green. Method 1 = OBS with 3 principal components – JADE with 35 components – correlation with mean squared BCG with 0.03; Method 2 = OBS with 1 principal component – JADE with 62 components – variance contribution 5%; Method 3 = OBS with 4 principal components – SOBI with 62 components – autocorrelation; Method 4 = OBS with 4 principal components – FastICA with 15 components - correlation with mean squared BCG with 0.08; Method 5 = OBS with 3 principal components – JADE with 35 components - correlation with mean BCG with 0.08.

Tables

Table 1: Effect of number of principal components (NPC) on the quality of the BCG removal with OBS evaluated with the first four validation criteria.

Validation criterion	**Effect of increasing NPC**	**Significance**
Peak-to-peak value	Strong decrease until NPC = 4 or 5	$p = 0.0002$
Single trial variation	Strong decrease throughout whole range of NPC	$p < 0.0001$
ERP difference	Weak decrease throughout whole range of NPC	$p = 0.0024$
Signal-to-noise ratio	No significant effect (highest SNR with NPC = 2)	$p = 0.4704$

The right column gives the p-value of the main effects from a repeated measurements ANOVA.

Table 2: Effect of number of principal components (NPC) on the quality of the BCG removal with OBS evaluated with task-related modulation values and classification percentages.

NPC	1	2	3	4	5	6	7
Detection (P1) [μV]	0.3801	1.3955	**1.3732**	**1.3051**	1.2452	1.2014	1.3154
Go/NoGo (N2) [μV]	**6.7355**	**6.8259**	**5.4281**	**5.4845**	**4.7942**	**4.8844**	**4.1677**
LVF stimuli [%]	**46**	**53.625**	**52.625**	**54.125**	**55.625**	**56.75**	53.375
RVF stimuli [%]	40.5	**54.25**	**54.875**	55.75	54.875	56.875	54

Upper part: Task-related modulation values (in microvolt) averaged over subjects for different numbers of principal components (NPC) with OBS. For the detection task the presented values are the averages over subjects of the differences between the contra- and ipsilateral amplitudes at the latency of the contralateral P1. For the Go/NoGo task the values represent the differences between the Go and NoGo amplitudes at the latency of the NoGo N2. Values in bold indicate a significant effect over all subjects.

Lower part: Average percentages of trials in the detection task in which the P1 amplitude difference between left and right corresponds with the side of stimulus presentation. Percentages are shown for both left and right visual field stimuli (LVF and RVF). Note that the shown percentages indicate the portion of trials that are really useful, meaning computed on the total number of trials in the experiment. Bold indicates values for which the amount of trials correctly classified between left and right is significantly higher than half of the number of trials after removal of extreme values (over subjects).

Table 3: Subset of good ICA parameters based on first four validation criteria.

Algorithm	**Number of components**	**Selection criteria**
JADE	All possibilities	- Correlation with ECG and mean BCG channel - Variance contribution - Peak-to-peak value of reconstructed sources
SOBI	All possibilities	- Correlation with ECG and mean BCG channel - Variance contribution
FastICA	50 and 62	- Correlation with ECG and mean BCG channel - Autocorrelation - Variance contribution - Peak-to-peak value of reconstructed sources
Infomax	25, 30, 35, 40, 50 and 62	- Correlation with ECG and mean BCG channel - Autocorrelation - Variance contribution - Peak-to-peak value of reconstructed sources

Possibilities of ICA parameters that remain after thresholding (based on minimal OBS values) on the first four validation criteria and after visual inspection of the grand average ERP.

Table 4: Evaluation of parameter possibilities from ICA subset with average percentages of trials of which the P1 amplitude difference between left and right corresponds with its stimulus type.

Algorithm	Number of components	Selection criterion	% LVF	p (LVF)	% RVF	p (RVF)
JADE	40	Correlation mean BCG 0.15	55.4	0.0299	59.0	0.0424
SOBI	40	Correlation ECG 0.15	51.6	0.0223	53.13	0.0278
Infomax	62	Correlation ECG 0.20	57	0.0256	59.25	0.0088
Infomax	62	Correlation mean BCG 0.20	55.13	0.0359	56.63	0.0369
Infomax	62	Peak-to-peak of reconstructed sources 25%	56.63	0.0115	61.25	0.0065

Percentages are shown for both left and right visual field stimuli (LVF and RVF). Only those parameter combinations are included for which the ratio accounting for real classification performance between left and right (i.e., after removal of extreme-value trials) is significantly > 50% over subjects (see p-values in columns 5 and 7).

Table 5: For one good parameter combination per algorithm from the ICA subset from Table 3 a comparison is made between OBS and ICA with the first four validation criteria.

Method (and parameters)	Peak-to-peak (DET)	Peak-to-peak (GNG)
OBS (NPC = 3)	7.76 %	7.82 %
FastICA (62 – PTP value of reconstructed sources 15%)	17.37 %	16.64 %
JADE (25 – correlation with mean BCG 0.15)	22.71 %	17.54 %
SOBI (35 – PTP value of reconstructed sources 15%)	7.55 %	8.03 %
Infomax (40 – variance contribution 5%)	34.53%	33.94%

Method (and parameters)	Single trial variation (DET)	Single trial variation (GNG)
OBS (NPC = 3)	57.02 %	53.82 %
FastICA (62 – PTP value of reconstructed sources 15%)	38.35 %	38.78 %
JADE (25 – correlation with mean BCG 0.15)	45.72 %	41.58 %
SOBI (35 – PTP value of reconstructed sources 15%)	21.95 %	22.73 %
Infomax (40 – variance	56.43%	56.13%

contribution 5%)				

Method (and parameters)	ERP diff central DET	ERP diff all DET	ERP diff Go	ERP diff NoGo
OBS (NPC = 3)	3.78 μV	2.83 μV	2.27 μV	2.62 μV
FastICA (62 – PTP value of reconstructed sources 15%)	3.05 μV	1.95 μV	2.07 μV	3.13 μV
JADE (25 – correlation with mean BCG 0.15)	3.18 μV	2.08 μV	2.20 μV	3.00 μV
SOBI (35 – PTP value of reconstructed sources 15%)	2.50 μV	1.45 μV	2.41 μV	3.50 μV
Infomax (40 – variance contribution 5%)	3.41 μV	2.34 μV	2.09 μV	2.85 μV

Method (and parameters)	SNR P1	SNR N1	SNR NoGo N2	SNR NoGo P3	SNR Go P3
OBS (NPC = 3)	6.75	12.27	8.08	13.70	25.50
FastICA (62 – PTP value of reconstructed sources 15%)	3.55	12.31	6.81	12.90	21.49
JADE (25 – correlation with mean BCG 0.15)	5.41	7.69	6.92	12.20	20.36
SOBI (35 – PTP value of reconstructed sources 15%)	2.34	13.97	6.93	8.95	22.79

Infomax (40 – variance contribution 5%)	5.92	10.08	9.83	16.27	21.70

The peak-to-peak value of the BCG epochs and the single trial variation are indicated in percent for both the detection (DET) and the Go/NoGo task (GNG). The ERP difference is given for the central stimulus and all stimuli together from the detection task (DET) and for both Go and NoGo ERPs from the Go/NoGo task (GNG). The SNR values are shown for the P1 and N1 peaks in the detection task and the NoGo N2 and P3 and the Go P3 in the Go/NoGo task.

Table 6: Subset of good OBS-ICA parameters based on SNR and task-related modulation.

Parameter	**Choice**	**Remarks**
Number of principal components	2-6	1 and 7 rarely give good results
Selection criterion	• Correlation with ECG • Correlation with mean channel • Correlation with mean squared channel (but not threshold 0.03) • Autocorrelation • Variance contribution	Using the frequency content does not remove any component after OBS
Algorithm and number of components	• JADE with at least 25 components • SOBI with at least 30 components • FastICA with 62 components • Infomax with at least 25 components	

Subset of parameter settings where OBS-ICA has an SNR not significantly worse than that of OBS (number of principal components = 3) and modulation values significant for both the detection and the Go/NoGo task.

Table 7: Evaluation of parameter possibilities from OBS-ICA subset (Table 7) with average percentages of trials of which the P1 amplitude difference between left and right hemisphere corresponds with its stimulus type.

NPC	Algorithm	Number of components	Selection criterion	% LVF	p (LVF)	% RVF	p (RVF)
3	/ (OBS alone)	/	/	52.63	/	54.88	/
2	SOBI	50	Correlation ECG, subject dependent	51.38	0.699	59.63	0.049
2	SOBI	62	Correlation mean, subject dependent	57.5	0.003	53.88	0.700
2	Infomax	62	Correlation mean, subject dependent	56.25	0.049	57.63	0.308
3	JADE	40	Correlation ECG, subject dependent	55	0.026	55	0.436
3	JADE	62	Variance contribution, 15%	50.63	0.769	56.63	0.022
3	SOBI	50	Correlation ECG, 0.15	52	0.746	57.8	0.032
3	Infomax	62	Correlation mean squared, 0.08	55.38	0.039	55.13	0.397
4	JADE	25	Correlation ECG, subject dependent	53.63	0.371	58.63	0.041

4	SOBI	40	Correlation mean, squared subject dependent	56	0.095	61.5	0.048
4	SOBI	62	Correlation ECG, 0.15	54.75	0.126	57.5	0.035
4	Infomax	30	Correlation mean, 0.03	54.88	0.184	58.25	0.011
4	Infomax	30	Correlation mean, 0.08	55.75	0.175	58.5	0.022
4	Infomax	30	Correlation mean, subject dependent	56.25	0.132	58.63	0.021
4	Infomax	40	Correlation mean, subject dependent	56.25	0.022	58.38	0.016
4	Infomax	50	Correlation mean, subject dependent	54	0.322	59.5	0.032
5	JADE	50	Correlation ECG, 0.07	57.25	0.034	57.25	0.017
5	SOBI	35	Correlation ECG, 0.07	52.75	0.446	58.5	0.024
5	Infomax	35	Correlation mean squared, subject dependent	51.88	0.629	59.63	0.031
5	Infomax	50	Correlation ECG,	59.25	0.077	59.75	0.042

			0.07				
5	Infomax	50	Correlation mean, subject dependent	52.88	0.46	60.63	0.015
6	JADE	50	Correlation mean squared, subject dependent	54	0.334	58.75	0.011
6	SOBI	50	Correlation mean squared, subject dependent	55.25	0.255	59.5	0.023
6	Infomax	50	Correlation mean, 0.03	55.13	0.197	58.88	0.019

Percentages are shown for both left and right visual field stimuli (LVF and RVF). P-values indicate whether the shown percentages are significantly better than the ones of pure OBS (with NPC = 3 – see first row in the table). Only those combinations with at least one of both p-values (from left and right stimuli) lower than 0.05 are retained here.