NeuroImage 59 (2012) 2607–2624

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Quantifying additive evoked contributions to the event-related potential

Georg Turi [a,b,c,*], Sascha Gotthardt [d], Wolf Singer [a], The Anh Vuong [e], Matthias Munk [f], Michael Wibral [b]

[a] Department of Neurophysiology, Max Planck Institute for Brain Research, Frankfurt, Germany
[b] MEG Unit, Brain Imaging Center, Goethe University, Frankfurt, Germany
[c] Cognitive Neuroscience Lab, Institute for Psychology, Goethe University, Frankfurt, Germany
[d] Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK
[e] Professur Graphische Datenverarbeitung, Institute for Informatics, Goethe University, Frankfurt, Germany
[f] Max Planck Institute for Biological Cybernetics, Tübingen, Germany

ARTICLE INFO

Article history:
Received 18 April 2011
Revised 25 July 2011
Accepted 25 August 2011
Available online 16 September 2011

Keywords:
Event-related potential
Local field potential
Electroencephalography
Phase-reset
Multiset independent component analysis
Generalized procrustes analysis

ABSTRACT

Event-related potentials (ERPs) are widely used in basic neuroscience and in clinical diagnostic procedures. In contrast, neurophysiological insights from ERPs have been limited, as several different mechanisms lead to ERPs. Apart from stereotypically repeated responses (additive evoked responses), these mechanisms are asymmetric amplitude modulations and phase-resetting of ongoing oscillatory activity. Therefore, a method is needed that differentiates between these mechanisms and moreover quantifies the stability of a response. We propose a constrained subspace independent component analysis that exploits the multivariate information present in the all-to-all relationship of recordings over trials. Our method identifies additive evoked activity and quantifies its stability over trials. We evaluate identification performance for biologically plausible simulation data and two neurophysiological test cases: Local field potential (LFP) recordings from a visuo-motor-integration task in the awake behaving macaque and magnetoencephalography (MEG) recordings of steady-state visual evoked fields (SSVEFs). In the LFPs we find additive evoked response contributions in visual areas V2/4 but not in primary motor cortex A4, although visually triggered ERPs were also observed in area A4. MEG-SSVEFs were mainly created by additive evoked response contributions. Our results demonstrate that the identification of additive evoked response contributions is possible both in invasive and in non-invasive electrophysiological recordings.

Introduction

Electrophysiological measurements such as invasive recordings of local field potentials (LFPs), electro- and magnetoencephalography (EEG/MEG), record signals from a multitude of neuronal sources. These recordings reflect various brain processes that may be related to some experimental stimulus (i.e. they are responses that are 'event-related' in a wide sense[1]), or may reflect background processes that are spontaneous and unrelated to the experiment. One of the earliest approaches to emphasize the responses of interest relative to background processes has been to average the recorded signals from multiple repetitions of the experiment (Galambos and Sheatz, 1962). The obtained average signal is usually called the event-related potential or field (ERP/F) for the case of electrical or magnetic measurements, respectively. The implicit assumption behind this approach was that there existed neuronal activity representing a response to the stimulus that would be *added* in a stereotypical repetitive fashion to each trial, i.e. phase-locked to the stimulus (additive evoked activity). Other activity that is event-related (in a wide sense) but not phase-locked to the stimulus (induced activity) or not event-related at all (spontaneous activity) should average to zero if enough trials were taken into consideration — resulting in an ERP/F that only contained the contributions of additive evoked activity (Bertrand and Tallon-Baudry, 2000; Galambos, 1992).

This assumption of additive evoked contributions has been challenged by the detection of two mechanisms that generate ERP/Fs without additive evoked response contributions (Fig. 1): First, the phase of spontaneous oscillatory activity could be altered (e.g. *reset*) by an event while its power remained unchanged, such that an ERP/F is observed, although one could not consider the contributions to be additive any more (Makeig et al., 2002; Sayers et al., 1974). Second,

* Corresponding author at: Department of Neurophysiology, Max Planck Institute for Brain Research, Deutschordenstrasse 46, 60528 Frankfurt, Germany; MEG Unit, Brain Imaging Center, Goethe University, Heinrich-Hoffmann-Strasse 10, 60528 Frankfurt, Germany; Cognitive Neuroscience Lab, Institute for Psychology, Goethe University, Frankfurt, Germany. Fax: +49 69 798 25209.
E-mail addresses: georg.turi@brain.mpg.de, turi@psych.uni-frankfurt.de, gturi@gmx.de (G. Turi).

[1] By using the term 'event-related in a wide sense' we try to distinguish this class of activity from the sub-class of so called 'evoked' activity that is for historic reasons also often referred to as 'event-related' activity. This is witnessed by the wide-spread use of the term 'event-related potential' when referring only to activity that survives the averaging over trials. Note that we take no fundamental standpoint in this issue, just try to define precise terms for the matter of this study.

1053-8119 © 2011 Elsevier Inc. Open access under CC BY-NC-ND license.
doi:10.1016/j.neuroimage.2011.08.078

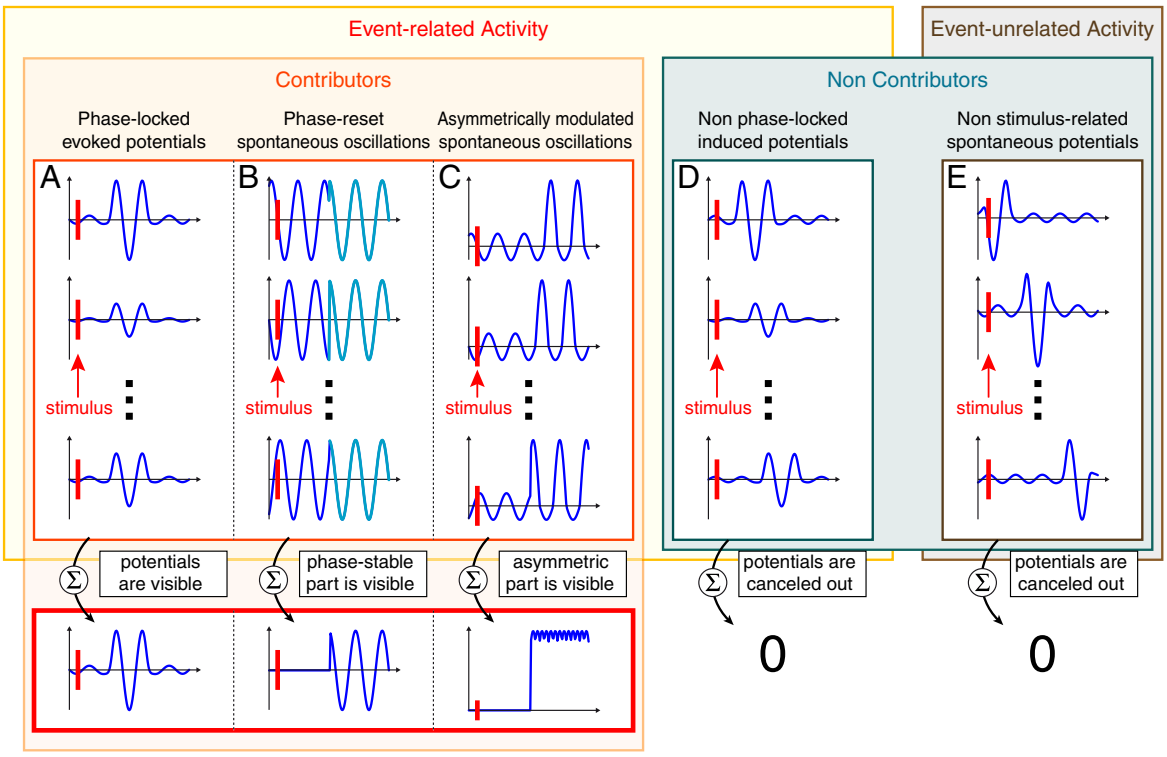


Fig. 1. Illustration of mechanisms leading to observable event-related potentials (ERPs). Neuronal activities may be related to an experimental stimulus (yellow box) or unrelated to the experiment (brown box). Stimulus-triggered trial averaging may elicit an observable ERP (orange box) or result in a vanishing signal (cyan box). (A) Additive evoked contributions are added stimulus-locked and phase-stable to the trials, i.e. the response potentials have a fixed polarity and latency with respect to the stimuli. The stereotypical repetitive response is conserved in the ERP. (B) Phase-reset spontaneous oscillations have varied phases in the pre-stimulus period over different trials. Following the stimulus, the phases of the oscillations are reset to a common value without changing the variances. Thus, oscillations in the pre-stimulus period are canceled out in the averaging process, leaving only the oscillations after the phase-reset in the ERP. (C) Asymmetrically amplitude modulated spontaneous oscillations have varied phases and symmetrically distributed amplitudes in the pre-stimulus period over different trials. Following the stimulus, the amplitudes of the oscillations are asymmetrically modulated around zero without changing the (random) phases. Thus, oscillations with symmetrically distributed amplitudes and varied phases in the pre-stimulus period are canceled out in the averaging process, leaving only the envelope of the amplitude changes in the ERP. (D) Additive induced contributions are added stimulus-related but phase-shifted to the trials, i.e. the response potentials have different polarities and shifted latencies with respect to the stimuli. The jittered responses are thus canceled out in the averaging process. (E) Spontaneous activity contributions are added non-stimulus-related and thus phase-varied to different trials. Hence, the potentials are canceled out in the averaging process.

an event-related amplitude modulation of spontaneous oscillatory activity that is *asymmetric* around zero could also result in a contribution to the ERP/F that will not average out, although the phase of these oscillations may be random. This theoretical prediction has recently been confirmed experimentally (Mazaheri and Jensen, 2008; Nikulin et al., 2007).

To separate the various generating mechanisms of the ERP/F from each other a variety of tests have been developed to find phase-reset (see Sauseng et al., 2007 for a critical review) and asymmetrically amplitude-modulated oscillations (Mazaheri and Jensen, 2008). Curiously, we still lack a test for the oldest of these hypotheses, the presence of additive evoked contributions to each trial. Their presence is typically inferred from negative evidence, i.e. by a failure to detect phase-resetting or asymmetric amplitude modulations.

Here we present an algorithm for quantifying additive evoked contributions to the ERP/F and demonstrate that these contributions can indeed be found in LFP recordings from awake behaving macaques performing a visuo-motor-integration task and in human MEG recordings of steady-state visual evoked fields (SSVEFs).

Methods

The remainder of this manuscript is organized as follows: In this section we will first present an overview over the various steps of the proposed algorithm, while referring the reader interested in an in-depth mathematical treatment to the appendix. We then describe tests of our algorithm on various sets of simulation data. Last, we apply our method to experimental datasets: (1) LFP data from the awake behaving macaque performing a visuo-motor-integration task where clear expectations on the presence of additive evoked contributions exist; (2) steady-state visual evoked field (SSVEF) MEG data, where additive response contributions are expected.

Overview of the algorithm

Here we present an outline of the proposed algorithm, highlighting its main concepts. We focus on the single channel version that is applicable even to invasive single channel recordings — the common multi-channel information in EEG/MEG may be used to further improve the method.

For the following description it is important to have a basic understanding of independent component analysis (ICA) (Comon, 1994). ICA describes a statistical method that can recover the unknown original source signals when only linear combinations (mixtures) of these source signals are observable and no details about this mixing are known. The only assumptions that are made in the ICA model are that there exist as many instantaneous (non-time delayed) mixtures as source signals and that the source signals are mutually statistically independent, while at most one source signal is Gaussian distributed (Hyvärinen and Oja, 2000). An example of such a mixture of source signals would be the observation of signals from multiple cortical sources on EEG electrodes because of volume conduction. These observed linear mixtures of the original source signals are usually referred to as *sensor signals*. The amplitude distribution of these

sensor signals will be closer to a Gaussian distribution than the amplitude distributions of the source signals (central limit theorem). If we observe as many sensor signals as we had source signals then for each source signal there would be a unique linear combination of the sensor signals that would result in this particular source signal again. Finding this linear combination is usually referred to as the 'demixing problem'. Sensor signals demixed by this linear combination would then have an amplitude distribution that would be further away from a Gaussian distribution (i.e. have a more peaked or flat distribution) than that of any other linear combination of sensor signals — because of the central limit theorem. ICA algorithms can be thought of as clever ways of finding these linear combinations. The estimated source signals are referred to as independent components (ICs).

This classical ICA model assumes as many sensor signals as there are source signals, while our goal is an algorithm that separates the contributions of additive evoked activity from contributions of other neuronal activity even in single-channel recordings. To make this problem tractable we reformulate the representation of our data in the following way: We start by noting that the single-channel signal of each trial is the addition of contributions of various neuronal activities that may elicit a non-vanishing average over trials (Fig. 1) and yet others that cancel out when enough trials are added. Contributions of additive evoked activity are special in three ways: First, they appear in every single-trial — as opposed to highly individual contributions of spontaneous activity to each trial. Second, they appear with a fixed polarity and latency to the given stimulus in each trial — in contrast to latency shifted contributions of induced activity to each trial. Third, they appear with an increased variance in the post-stimulus period of each trial — in contrast to the variance stable contributions of phase-reset activity in each trial.

To exploit these three properties we apply the following central idea as the first step for our algorithm: We take the single-trial signals and treat them *as if* they were all recorded by different channels at the same time, i.e. we represent every trial of the single-channel dataset as a separate channel of a multichannel dataset having only one trial. These channels of the new multichannel dataset can now be treated as the sensors of the classical ICA model (Figs. 2, A, B). This way, each sensor signal of the ICA model corresponds to one original single-trial signal. To reflect this definition we will refer to single-trials as 'sensors' in the remainder of the text and each time we do use the word *sensor* in the following it will explicitly mean a single-trial signal of the original single-channel recording.

In the next four paragraphs we will explain how the contributions of each of the four different activities (additive evoked, spontaneous, induced and phase-reset) will appear as mixtures of source signals in the classical ICA model when we use trials as sensors.

On our new sensors, created from the single-trials, additive evoked contributions (AECs) appear with the same polarity and latency on each sensor — remember the phase-locking property of additive evoked activity. This property is required for the identifiability of source signals in the classical ICA mixing model (see above). Therefore, additive evoked activity generated by *one* group of neurons can be considered as *one* source signal that is instantaneously mixed onto all sensors (see Fig. 1 first column).

The phase-locking property does obviously not hold for spontaneous contributions that do not repeat over trials. Although the spontaneous activity may be generated by the same group of neurons in each trial the temporal activity of these neurons may well be statistically independent over trials. Thus, spontaneous activity generated by this single group of neurons will be seen as a set of several distinct source signals when transforming trials into sensors (Fig. 1, fifth column). Each of these source signals will then be mixed onto one sensor only. In the presence of several groups of neurons generating spontaneous activity we will therefore have more sources than sensors — in principle this is the overcomplete bases problem for ICA (Lewicki and Sejnowski, 2000). However, since the spontaneous contribution is individual to each sensor but has a similar amplitude distribution on each sensor it can be considered as (uncorrelated) sensor noise.

The phase-locking property does by definition also not hold for induced contributions. This is because induced contributions have shifted latencies (phases) over the sensors. Here the contributions of induced activity generated by a single group of neurons are neither fully statistically independent over sensors nor are they identical as it was the case for AECs. Thus, induced activity generated by *one* group of neurons can be considered as *one* source signal that is mixed in a time-delayed way onto the sensors (see Fig. 1 fourth column). In the classical ICA model, such source signals cannot be recovered at all (see e.g. Hyvärinen et al., 2001). This is because time-delayed mixtures of source signals will be seen as sparse mixtures of more source signals than sensor signals when transforming trials into sensors (see Appendix). Therefore, we can consider induced contributions as (correlated) sensor noise for the same reason as given above (i.e. individual contributions to each sensor).

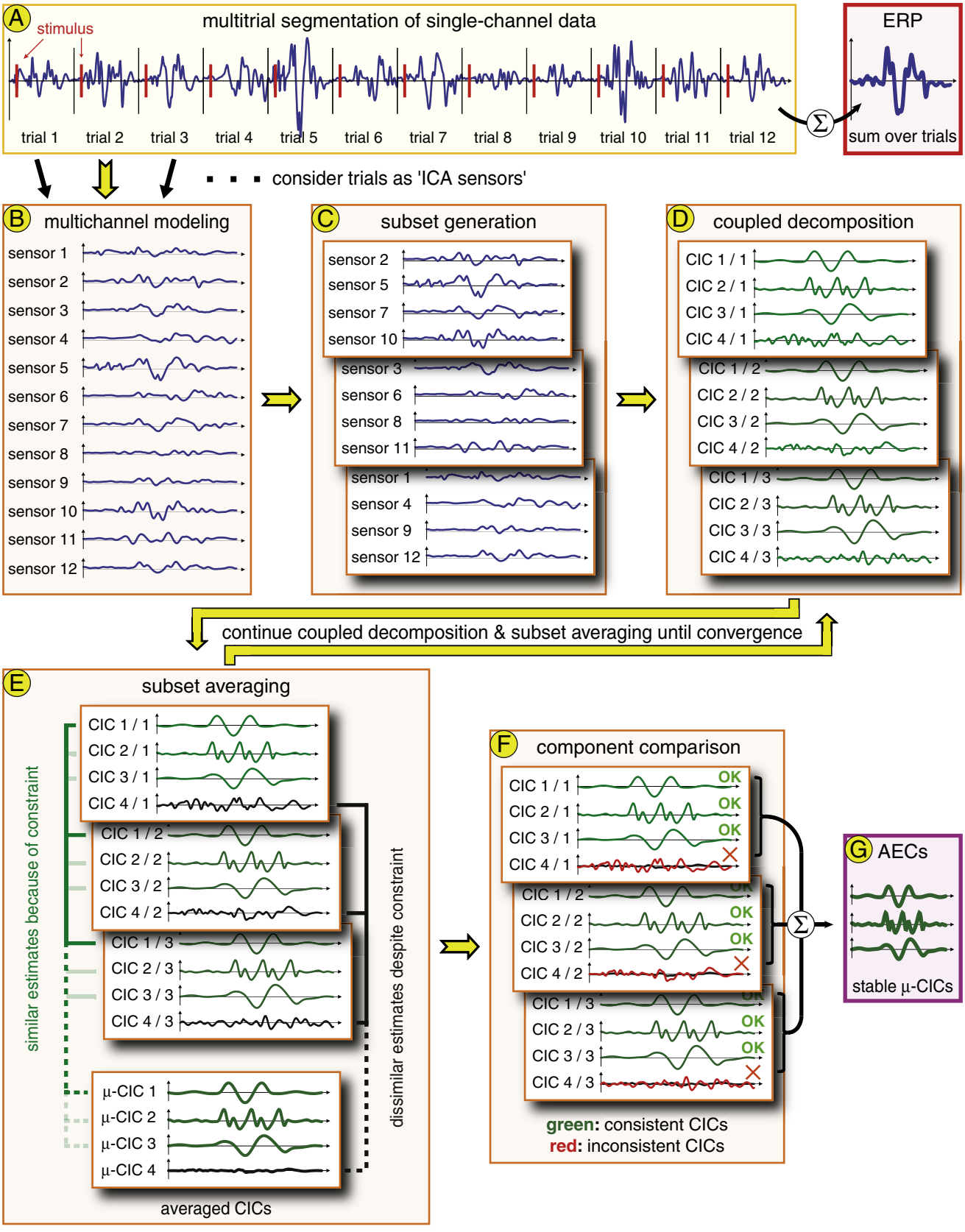
The phase-locking property also takes an exceptional form for phase-reset contributions. This is because phase-reset contributions have a fixed latency only in the post-stimulus period. Here the contributions of phase-reset activity generated by a single group of neurons are neither completely shifted over the sensors nor are they fully statistically independent as it was the case for spontaneous contributions. Thus, phase-reset activity generated by *one* group of neurons can be considered as *one* source signal that is mixed in a partially time-delayed way onto the sensors (Fig. 1 second column). However, the classical ICA model does not account for phase-reset sources for the same reasons as for induced and spontaneous source signals — phase-reset sources have an *individual* contribution to each sensor (single-trial signal) because of their distinct pre-stimulus (non phase-reset) part, which must be included in the analysis. Therefore, we can consider phase-reset contributions as (correlated) sensor noise for the same reason as given above (individual contributions to each sensor).

At this point our ICA mixing model comprises additive evoked sources contributing to every sensor, spontaneous sources contributing only to single sensors, induced sources that contribute in a time-delayed way to the sensors and phase-reset sources that contribute in a partially time-delayed way to the sensors. The latter three reflect sources contributing only to single sensors and therefore we consider them as sensor noise.

This new formulation of the problem lets us see the following important property: The fact that additive evoked sources contribute instantaneously to every sensor means that we could choose any subset of our sensors and still *all* additive evoked sources would contribute to our subset. However, any given subset of sensors will be free of spontaneous, induced and phase-reset contributions that are specific to other sensors, i.e. those not considered for the particular subset. We will now describe how we use this property of subsets to attenuate spontaneous, induced and phase-reset contributions (i.e. the noise) and recover the AECs only.

Based on the idea above, we choose several non-identical subsets of an equal number of sensors (Fig. 2, C). We assume that the number of sensors in each subset is higher than the number of AECs. Our goal is to decompose the data from each subset in such a way that each of the AECs is represented as one estimated source signal — perhaps with some distortions (see Appendix). The contributions of all other sources should be represented by one or more additional nuisance signals, which should be clearly separable from AECs. To this end, we consider the problem as a constrained subspace ICA where only the mutually independent AECs are considered as the source signals to be recovered. All other contributions are considered as noise to be attenuated and not recovered individually. An approximate solution to our constrained ICA problem is to simultaneously decompose all subsets in a coupled way, by maximizing the mutual independence of the estimated components of each single subset decomposition

and using a similarity constraint that concurrently forces the decompositions to converge to such signal components that are most similar over all subsets (Fig. 2, D). These criteria are equivalent to our goal stated above because the mutually independent AECs satisfy both criteria, i.e. mutual independence within each subset and identity over all subsets and will therefore be recovered by the coupled decomposition. In this coupled decomposition the number of sensors in each subset needs to be larger than the number of AECs since the



sensor signals are mixtures of additive evoked and the other contributions. This is because we need the possibility to estimate additional source signals reflecting the residual contributions of induced, spontaneous and phase-reset sources that are specific to each subset. This way, some of the estimated source signals will approximate independent components (ICs) that are *common* to all subsets — i.e. consistently present in all subsets. Therefore, we will refer to such consistent ICs as *common* ICs (CICs). However, since we do not have *a priori* knowledge about the number of truly consistent CICs in our data, we will preliminarily consider all CICs estimated by our algorithm to be consistent. Later on, we will describe how we distinguish consistent CICs from inconsistent ones.

The main difficulty to overcome is the design of a similarity constraint for these coupled simultaneous subset decompositions. To implement the constraint we make use of the fact that for each CIC there is only one unique mixing coefficient for each sensor. However, of this coefficient we obtain multiple estimates by decomposing multiple subsets that contain that particular sensor. An elegant way to enforce convergence to a global optimum is to constrain the estimated mixing coefficients corresponding to the *same* CICs and sensors to converge to the *same* values. For this, we first perform a single iteration step of a fixed-point ICA (Hyvärinen, 1999a) decomposition for each subset. Then, we compute a *global* mixing matrix for *all* CICs and sensors using the estimated coefficients from the subset decompositions. This global mixing matrix will typically be rectangular, with the coefficients in the rows corresponding to the sensors and the coefficients in the columns corresponding to the CICs. One estimate of the coefficients for certain sensors (rows) and all CICs (columns) is provided by each single subset decomposition. Due to the fact that we have many more possible subsets than we have single sensors, we get — as outlined above — repeated estimates of the coefficients in the global mixing matrix. To get an estimate of the unique mixing coefficient for a given CIC and sensor combination we therefore average over all corresponding estimates obtained from the subset decomposition. This average is then used as a starting point for the reestimation of the subset mixing coefficients in the next iteration step of the subset decompositions. The algorithm then iteratively alternates between one step of subset decompositions and one step of coefficient reestimations until all mixing coefficients become stable — i.e. the differences between corresponding mixing coefficients from the subset decompositions and the global mixing matrix are negligible (see Appendix).

The convergence to stable mixing coefficients is further improved by the similarity constraint for the CICs themselves: So far we have only used the fact that there exist unique mixing coefficients for all CIC and sensor combinations. However, we also know that unique CICs exist. Therefore corresponding CICs over the subsets should be as similar as possible — at least for independent AECs. We can implement this knowledge by changing the subset mixing coefficients *before* the coefficient averaging step in such a way that corresponding CICs over the subsets become more similar. To this end we average over all corresponding CICs and modify the subset mixing coefficients to maximize the non-linear cross-correlation between CICs and corresponding mean CICs (μ-CICs) (Fig. 2, E). This procedure will also increase the similarity between all corresponding CICs over the subsets if the variances of the CICs are kept fixed during the iterations (see Appendix). The alternating steps of subset decompositions, similarity maximizations and coefficient stabilizations are then iterated until the algorithm converges.

After the convergence of the algorithm we have to judge whether the obtained CICs are consistent over the subsets and thus can be accepted as AECs. To this end we use a similarity measure that puts the focus on the actual presence (i.e. consistence) of AECs. In our mixing model such a similarity measure is given by the 'stability' of the subset mixing coefficients — quantified as the mean over the normalized inner-products of corresponding subset and global mixing coefficients (Fig. 2, F). This way, the obtained stability value will gradually indicate the presence of an AEC in the range of [0, 1]. We will refer to the stability value as Θ and the decision parameter for critical stability as Θ_c in the remainder of the text. Consequently, contributions being consistent over the subsets (AECs) will have stability indices Θ close to 1, while inconsistent contributions (noise residuals) will have stability indices clearly lower than 1.

Last, we note that if only very little pre-stimulus data are included in the analysis, then mean CICs (μ-CICs) for phase-reset activity might be linked to subset mixing coefficients with high stability indices as well and therefore might be misclassified as AECs. This is because the post-stimulus period of phase-reset activity is by definition phase-stable and thus identical over trials. Unfortunately, μ-CICs representing phase-reset contributions might also have a pre-stimulus period that is close to zero due to the pre-stimulus random phases of phase-reset activity. Nevertheless, a criterion to distinguish phase-reset from AECs can be derived based on the following observation: CICs representing phase-reset contributions obtained from single subset decompositions effectively reflect a weighted average (linear combination) over a smaller number of sensors than the final μ-CICs. The CICs representing phase-reset contributions therefore have an increased variance in the pre-stimulus period already, compared to their corresponding μ-CICs. This is different from AECs that have — due to their additive property — a pre-stimulus period close to zero both in the μ-CIC and in each CIC obtained from a subset decomposition. Hence, we can argue that if the CICs from the subsets have only a relatively small difference in variance between pre- and post-stimulus period they can be suspected to reflect phase-reset contributions. In contrast, if a CIC from a subset has a large increase in variance between pre- and post-stimulus period it very likely represents an AEC (see Fig. 3 for examples of both cases). The expected relative increase η_e between pre- and post-stimulus variance due to the cancelation of random phases φ_j (i.e. for phase-reset contributions) in the pre-stimulus period over a number of N trials (the size of the subset) is:

$$\eta_e = \frac{Power_{phase-reset} - Power_{random-phase}}{Power_{random-phase}} = \frac{1 - \frac{1}{N}\sum_{j=1}^{N} \exp\left(i\varphi_j\right)}{\frac{1}{N}\sum_{j=1}^{N} \exp\left(i\varphi_j\right)} = \frac{1}{\frac{1}{N}\sum_{j=1}^{N} \exp\left(i\varphi_j\right)} - 1$$

Fig. 2. Illustration of the Common Independent Component Analysis (CICA) algorithm. (A) Segmentation of a multitrial single-channel recording. Left: Segmentation into equally sized single-trial signals triggered to the stimulus onset (red bars). Right: ERP calculated by averaging over trials. (B) Modeling of a multichannel system. Each single-trial signal is treated as an individual sensor signal of a simultaneous multichannel recording. (C) Generation of signal subsets. The multichannel system is subdivided into subsets, each consisting of a same number of randomly chosen sensor signals. Subsets have to differ in at least one sensor index. (D) Coupled decomposition of the subsets into common independent components (CICs). Estimated CICs are constrained to be maximally, mutually independent in each subset and maximally similar to CICs with corresponding indices over all subsets. (E) Averaging of the decomposed subsets. Top: Green traces in the subsets depict similar CICs because of the constraint (see D), black traces depict dissimilar CICs despite the constraint. Bottom: Mean CICs (μ-CICs) obtained by averaging of index equal CICs over the subsets. Bold green traces: μ-CICs composed of similar CICs, bold black traces: μ-CIC composed of dissimilar CICs. The μ-CICs are used as a constraint for the next round of coupled decomposition. The alternation of coupled decomposition and mean estimation is continued until convergence. (F) Comparison of CICs and corresponding μ-CICs. Consistent CICs (green traces) are similar to corresponding μ-CICs and thus have stable mixing coefficients (identical green traces), while inconsistent CICs (red traces) are dissimilar to corresponding μ-CICs and have instable mixing coefficients (black traces). (G) Selection of additive evoked contributions (AECs). Consistent CICs are accepted as AECs if their mixing coefficients satisfy a given stability criterion. Identified AECs are finally represented by corresponding 'stable' μ-CICs.

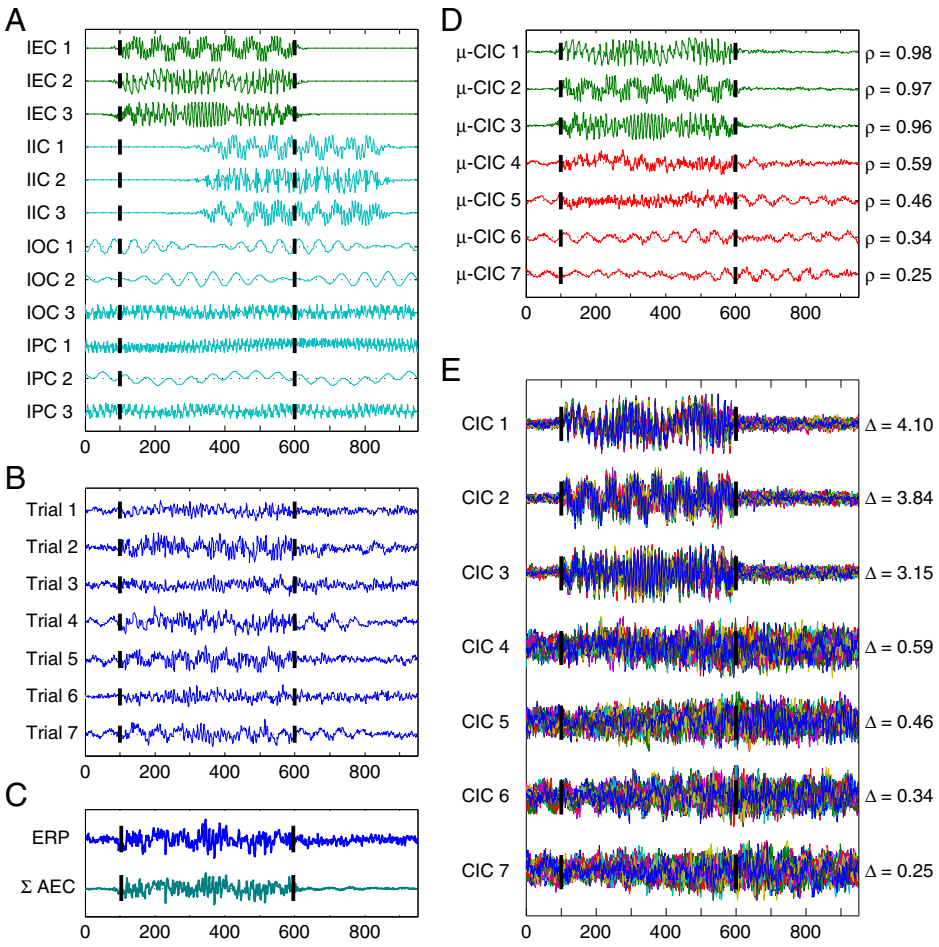


Fig. 3. Example of identifying simulated AECs by estimating CICs (central 950 of a total of 1500 simulated points shown). Black bars indicate begin and end of the designated post-stimulus task periods. (A) Simulated independent evoked components (IEC 1–3), independent induced components (IIC 1–3), independent 'ongoing' spontaneous components (IOC 1–3) and independent phase-reset components (IPC 1–3). In the case of IOCs only 3 of 300 components are shown. Mixing onto the trials was performed non-time-delayed for IECs, time-delayed with different latencies for IICs, partially time-delayed with fixed latencies in the designated task period for IPCs and completely sparse for IOCs. (B) Seven example trial signals (of a total of 100). (C) ERPs obtained from averaging 100 trial signals (ERP) and the sum over the estimated additive evoked contributions to 100 trials (Σ AEC). (D) Results obtained when estimating 7 CICs from the trial data, ranked according their stability indices in descending order. (Green, identified AECs = μ-CICs reflecting consistent CICs that satisfy the stability criterion of $\Theta_c = 0.97$ and variance criterion of $\Delta_c = 1$; red, noise residuals = μ-CICs reflecting inconsistent CICs). The linear cross-correlation coefficients of μ-CICs and IECs are depicted on the right. The number of identified AECs (green traces) equals the number of simulated IECs. (E) Superposition of the results obtained from the corresponding subset decompositions. The variance indices of the estimated CICs are depicted on the right. Note the significant variance increase in the designated task period of the first 3 CICs (recovered AECs).

Therefore, we suggest using an average variance increase $\eta/\eta_e > 1$ in the CICs as an additional criterion to test for AECs. We will refer to this important decision ratio as the critical variance index in the remainder of the text and denote it by the symbol Δ_c from here on.

To understand the above argument it may be helpful to think of how a dataset containing only one AEC and no noise is being averaged over subsets of sensors (i.e. trials) and finally over all sensors — one will always obtain an average that has a zero baseline. In contrast, a dataset containing only one phase-reset component and no noise will result in sensor subset averages that have a non-zero baseline whereas the final average will have a close to zero baseline. In this phase-reset dataset we would observe a difference in baseline amplitude depending on the number of sensors (i.e. trials) we use for averaging. The relation to the above argument becomes clear if we recall that a CIC reflects a subset average of a component (AEC or phase-reset) and the μ-CIC the full average.

Performance evaluation on simulation data

We tested our algorithm on two groups of datasets. The first group of datasets was used to test the identification of AECs on mutually independent source signals mixed by our mixing model. Thus, contributions of additive evoked, induced, phase-reset and spontaneous activities were approximated by distinct sets of mutually independent random variables that were mixed non-time-delayed, time-delayed, partially time-delayed and sparse onto the simulated single-trials. The second group of datasets was used to test the identification of AECs in the case of temporal correlations and dominant variance increases reflecting biologically plausible constraints. To this aim the source signals were not fully independent because of their temporal structure as explained below (also see Fig. 3 for examples). For the quantification of the performance we evaluated the algorithm on 500 different datasets of 100 simulated single-trials each.

For each of the datasets in the first group, we initially created 30 ICs by sampling 500 data points from 15 distinct sub-Gaussian and 15 distinct super-Gaussian random variables. Each of the ICs was mean-free and had unit variance. Next, we created 100 linear mixtures of 10 of the ICs (5 sub- and 5 super-Gaussian) using uniformly distributed, positive mixing coefficients. According to our mixing model, we considered these mixtures as contributions of 10 independent additive evoked sources. Additionally, we created 100 time-delayed linear mixtures of the remaining 20 ICs using uniformly

distributed, mean-free mixing coefficients. Of these 20 ICs, 10 ICs (5 sub- and 5 super-Gaussian) were shifted over the entire signal length, while the remaining 10 ICs were only shifted in the first and last third of the signal. The corresponding shifting factors were randomly chosen from an interval of 10% of the signal length, i.e. [1, 2, …, 50] data points. We considered these mixtures as contributions of 10 independent induced and 10 independent phase-reset sources, respectively. In order to obtain contributions of independent spontaneous sources we created 100 sparse linear mixtures of 1000 distinct ICs using mean-free, uniformly distributed mixing coefficients. Each mixture consisted of the contributions of 10 different ICs that were created by sampling 500 data points from 5 sub-Gaussian and 5 super-Gaussian random variables. All ICs were mean-free and had unit variance. Finally, we created trial signals by summing up the sets of mixtures, removing the obtained signal means and normalizing their variances to unity.

For each of the datasets of the second group, we initially created 3 small sets of 3 distinct sinusoidal signals and 1 large set of 300 distinct sinusoidal signals. Each signal had a length of 1500 data points and a frequency randomly chosen from an interval of [1, 2, …, 100] Hz. Next, we cross-modulated the amplitude and frequency of the signals by the frequencies of the other signals in the same set. This way we introduced temporal linear correlations within and statistical dependencies among the signals in a set. In two of the smaller sets, we then set the values of the first and last third of the signals to zero and imposed this way a variance increase onto 500 data points in the middle of these signals. In order to obtain *maximally* mutually independent signals, we subsequently removed the signal means, normalized their variances to unity and performed an ICA separately on each set using *EFICA* (Koldovsky et al., 2006). On the large set of 300 sinusoidal signals, we performed ICA in a block wise manner on 3 consecutive signals at a time. Hereafter the obtained ICs were maximally mutually independent but still had temporal features and statistical dependencies that are characteristic for stimulus responses recorded in vivo. Next, we created 100 linear mixtures and 100 time-delayed linear mixtures of the first and second IC set having a variance increase, respectively. Additionally, we created 100 partially time-delayed and 100 sparse mixtures of the remaining smaller and the large IC set, respectively. We considered the mixtures having a variance increase as contributions from independent evoked and independent induced sources, and the mixtures without a variance increase as contributions of independent phase-reset and independent spontaneous sources. Parameters for these mixings were chosen as for the first group of datasets. Again, shifting factors for time-delayed contributions were randomly chosen from 10% of the signal length and independent phase-reset contributions were shifted only in the first and last 500 data points of the signals. Furthermore, each sparse mixture of independent spontaneous contributions consisted of the contributions of 3 different ICs. Finally, we created trial signals by summing up the sets of mixtures, removing the obtained signal means and normalizing their variances to unity.

We quantified the performance of our algorithm by evaluating true-positive and false-positive rates as follows: A μ-CIC was counted as a true-positive identification of an AEC if it had a linear cross-correlation with the original independent AEC that exceeded 0.9 and additionally if the mixing coefficients of the corresponding CICs were stable over the subsets and their post-stimulus variance was significantly increased — i.e. by exceeding the critical threshold Θ_c for the stability index and the critical threshold Δ_c for the variance index. In contrast, we counted a μ-CIC as a false-positive identification if it had a linear cross-correlation with the original independent AEC below 0.9, but both its mixing coefficients were stable over the subsets as well as its post-stimulus variance was significantly increased.

As the identification of an AEC depends on the chosen critical stability index Θ_c we quantified the performance over a wide interval of thresholds for the critical stability index [0.8, 0.81, …, 0.99] and also a number of estimated CICs [11, 12, …, 20] and [4, 5, …, 10] for each dataset of the first and second group, respectively. For the datasets of the second group we used a fixed threshold for the critical variance index of $\Delta_c = 1$.

Performance on LFP – VEP and MEG – SSVEF data

Simulation data may always differ from biological signals in some hidden quality. This difference may be small but still could favor good performance of our algorithm. To fully test the algorithm we therefore need biological data where we can expect AECs to be present – or absent – in the data. Unfortunately in biological data the ground truth is unknown. We therefore make use of the following working hypothesis to construct a first neurophysiological test case: If the data were recorded only one or very few synapses away from sensory inputs, chances are relatively high that responses are stereotypically repeated over trials. We assume that in this case one can observe AECs to the ERP. In contrast, when data are recorded several synaptic connections away from the sensory input we do not expect AECs to be present. This is because intrinsic activity of the neural networks between sensory input and recorded sites will exert a variable influence on the response such that it either becomes jittered in time or has a varying wave shape in each trial.

Here, we used LFP data recorded transcortically using bipolar electrodes (as in Bressler et al., 1993) in various cortical areas (visual areas: V2, V4 and primary motor cortex: Brodmann area A4) of a macaque monkey (Fig. 5). All procedures were approved by the local authorities (Regierungspräsidium) and are in full compliance with the guidelines of the European Community (EUVD 86/609/EEC). Data were recorded at a sampling rate of 1000 Hz using a hardware band-pass filter with a high cut-off frequency of 150 Hz and a low cut-off frequency of 5 Hz (3 dB/octave). In each recording session, the monkey performed one of two visuo-motor-integration tasks: The monkey was required to fixate a spot at the center of a CRT screen. In task 1, after the monkey kept fixation for 300–1000 ms, a drifting sinusoidal grating appeared contralateral to the recorded hemisphere. In task 2, after the initial fixation period of 300–1000 ms, two drifting gratings appeared — one at each side of the fixation point. These stimuli were presented in parafoveal positions to match electrode positions in visual areas. The color of the fixation point cued the task relevant grating where attention had to be directed (green: left, red: right). This way the grating contralateral of the recording site was either attended or unattended.

In both tasks the monkey had to maintain fixation of the central fixation spot and indicate the varying speed of the task relevant grating by moving a manipulandum lever against the force of springs to predefined positions. For the analysis the data were cut into 512 ms long parts from 100 ms prior to stimulus onset to 412 ms post-stimulus onset. As explained above, our expectation was to identify AECs in visual cortices and we did not expect AECs in recordings from primary motor cortex A4, although motor cortex neurons may indeed respond to visual stimulation (e.g. Merchant et al., 2001). We used data from three different sessions to obtain recordings with optimal signal-to-noise ratio: Data for the analysis of ERPs from visual area V4 and primary motor area A4 were taken from sessions when the monkey performed task 1, data for the analysis of ERPs from visual area V2 were taken from a session when the monkey performed task 2. Only trials, in which the monkey correctly performed the task and, in the latter case, attended the grating contralateral to the recording site, were used.

The second test case was constructed using visual evoked responses to regularly repeating stimuli — steady-state visual evoked potentials or fields (SSVEP/Fs). These SSVEFs are known to accurately follow the repetition frequency of the stimulus (e.g. Regan, 1977). In addition, the Fourier transform of such steady-state potential signals exhibits power at multiple harmonics of the stimulus repetition

frequency. This means that the wave shape of the response is highly conserved from stimulus presentation to stimulus presentation and from stimulus train to stimulus train. Therefore, we expect AECs to be present in these responses. We tested this hypothesis using data from a single-channel (MRO22) recorded at a sampling rate of 1.2 kHz with a whole head magnetoencephalograph (VSM MedTech Ltd., Canada) while the subject passively viewed an array of LEDs in the lower left visual hemifield. In each trial LEDs flickered at rates of either 10, 21, 42.1 or 84.2 Hz for a stimulus train length of 3.6 s, with a preceding fixation interval (baseline) of 3.6 s. 60 trials per condition were recorded. Here, we present only results for the 21 Hz condition. Data were filtered between 0.5 and 150 Hz using Fieldtrip (v20081210). Trials were defined from −500 to 3600 ms with respect to stimulus onset and the mean of the pre-stimulus baseline period was subtracted. Trials contaminated by artifacts (eye blinks and movements, muscular activity) were rejected using the automated artifact rejection routines in Fieldtrip. We then extracted the signal of the sensor showing highest power in the stimulus frequency for further analysis. For comparison, we also computed the event related spectral perturbation and the inter-trial phase-locking values for the same data using EEGLAB (Delorme and Makeig, 2004).

Results

Results from the performance evaluation of our algorithm on the first group of datasets indicated an overall high sensitivity and specificity to the identification of simulated mutually independent additive evoked contributions (AECs) (see Fig. 4). Evaluation with a fixed number of estimated CICs showed a progressive decrease of the false-positive rates and a general conservation of high true-positive rates when increasing the critical stability index Θ_c (see Fig. 4, A). This decrease of the false-positive rate was characteristic for all estimation runs using a higher number of estimated CICs than the number of simulated AECs in the data (see red curves in Figs. 4, A and B). In particular, no estimation run showed a false-positive identification when using $\Theta_c > 0.96$, while the true-positive rate was above 0.95 when estimating more than 12 CICs (i.e. more than 2 additional CICs accounting for the noise residuals) using $\Theta_c = 0.96$. Furthermore, the margin of separation between *high* true-positive and *minimum* (zero) false-positive rates was largest when using a critical stability index of $\Theta_c = 0.97$ (see vertical black dotted line in Fig. 4, A). In this case, the true-positive rates were generally above 0.99 if the number of estimated CICs was at least 1.5

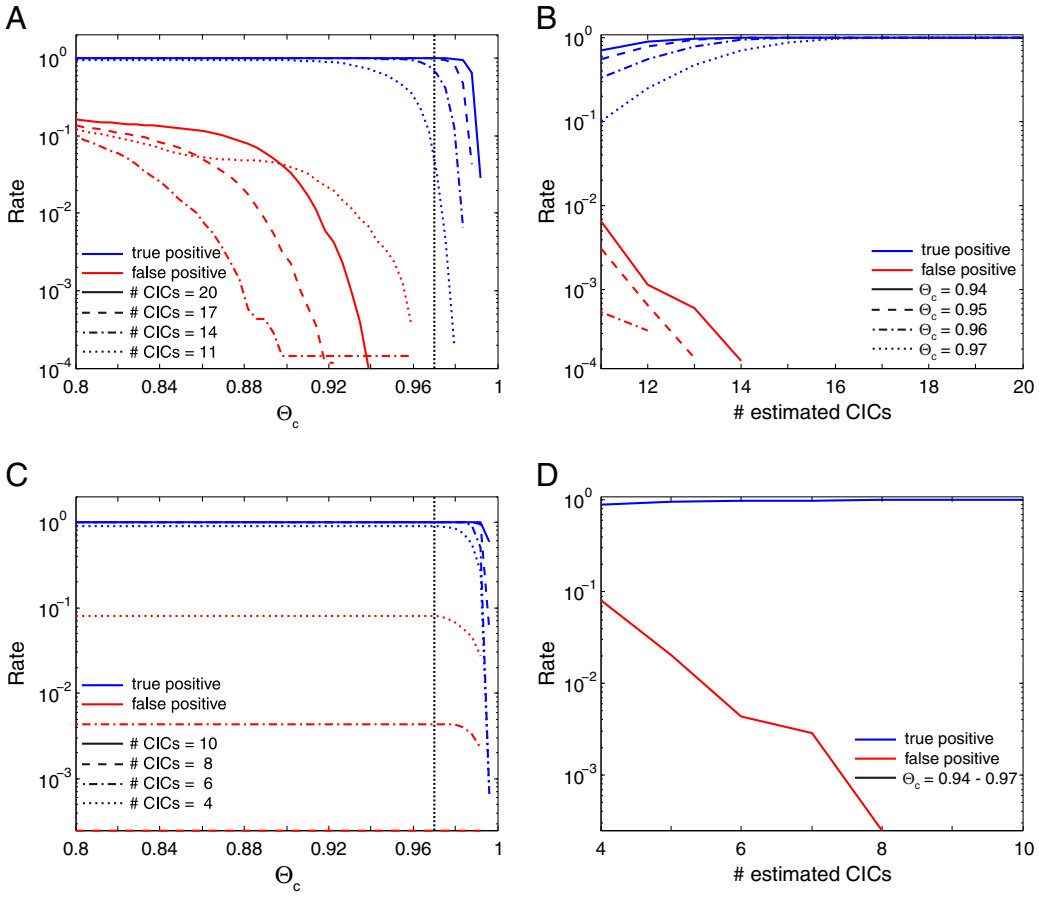


Fig. 4. Performance curves on mutually independent simulation data (top part) and biologically plausible simulation data (bottom part) showing average true (blue) and false (red) positive identification rates (500 repeated estimations) on a logarithmic scale. Common independent components (CICs) were accepted as recovered AECs if the corresponding stability index Θ exceeded the chosen critical value Θ_c and for the case of the biological plausible data the corresponding variance index Δ also exceeded the critical value $\Delta_c = 1$. Recovered AECs were counted as true-positive identifications, if the correlation coefficient between the corresponding mean CIC (μ-CIC) and the original simulated AEC was higher than 0.9 — otherwise they were counted as false-positive. *Top part*: (A) Dependence of the performance on Θ_c for different subset sizes: 20 (solid lines), 17 (dashed lines), 14 (dash-dotted lines), 11 (dotted lines). The vertical dotted line corresponds to $\Theta_c = 0.97$. (B) Dependence on the subset size for different Θ_c: 0.94 (solid lines), 0.95 (dashed lines), 0.96 (dash-dotted lines), 0.97 (dotted lines). No false-positives were generated for $\Theta_c > 0.96$ or a subset size larger than 14. *Bottom part*: (C) Dependence of the performance on Θ_c for different subset sizes: 10 (solid lines), 8 (dashed lines), 6 (dash-dotted lines), 4 (dotted lines). The vertical dotted line corresponds to $\Theta_c = 0.97$. (D) Dependence on the subset size for different Θ_c: 0.94–0.97 (solid lines). All tested Θ_c resulted in similar results, while no false-positives were generated for a subset size larger than 8.

times higher than the number of simulated AECs in the data (see blue curves in Fig. 4, B).

Similar results were obtained from the performance evaluation on the second set of simulation data. Despite the fact that the simulated biologically plausible AECs were not completely mutually independent, the results indicated a high sensitivity and specificity to the identification of AECs with temporal correlations and dominant variance increases as well. Evaluation with a fixed number of

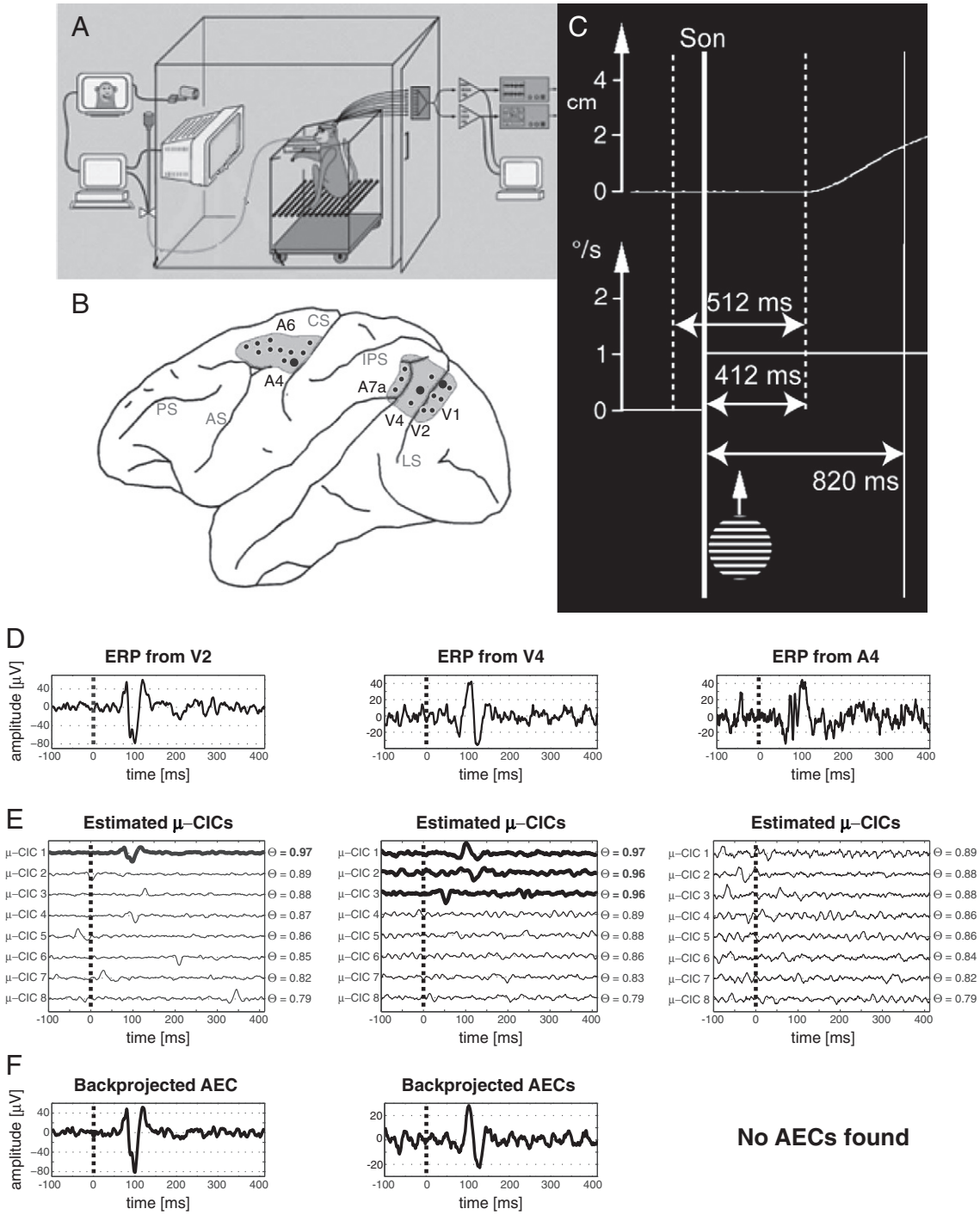


Fig. 5. (A) Experimental setup of *in vivo* LFP recordings from a macaque monkey performing a visuo-motor task. The monkey was placed in a primate chair fixating the center of a CRT screen and had to move a lever to predefined positions corresponding to the direction and speed of drifting gratings presented on the screen. (B) Left hemisphere of a monkey brain with all recording sites (gray shading) of the LFPs in visual, parietal, motor and premotor areas. The LFPs analyzed in this study were recorded at the sites marked in bold. (C) Example of a lever trajectory from a single-trial (top part) and the corresponding direction and speed (degree/s) of the presented drifting grating (bottom part). Bold vertical line (Son): stimulus onset. Dotted vertical lines: analysis window from 100 ms pre-stimulus to 412 ms post-stimulus. Regular vertical line, to the right: arrival time at the predefined lever position. (D) Event-related potentials (ERPs) of LFP recordings from areas V2, V4 and A4: time zero corresponds to the onset of the visual stimulus (vertical dotted lines). (E) Analysis of additive evoked contributions (AECs) for the above ERPs. AECs were accepted as identified if the corresponding mixing coefficients of the estimated common independent components (CICs) exceeded a critical stability index of $\Theta_c = 0.95$. AECs were identified in visual areas V2 (1 AEC) and V4 (3 AECs) and are indicated by bold lines. Remaining μ-CICs indicated by regular lines. No AEC was identified in primary motor area A4. (F) Average over trials of the 'backprojected' AECs identified by the algorithm.

estimated CICs showed a progressive decrease of the false-positive rates and a general conservation of high true-positive rates when increasing the critical stability index Θ_c (see Fig. 4, C). Furthermore, all estimation runs showed a characteristic decrease of the false-positive rate when using a higher number of estimated CICs than the number of simulated AECs in the data. This decrease, however, was similar for

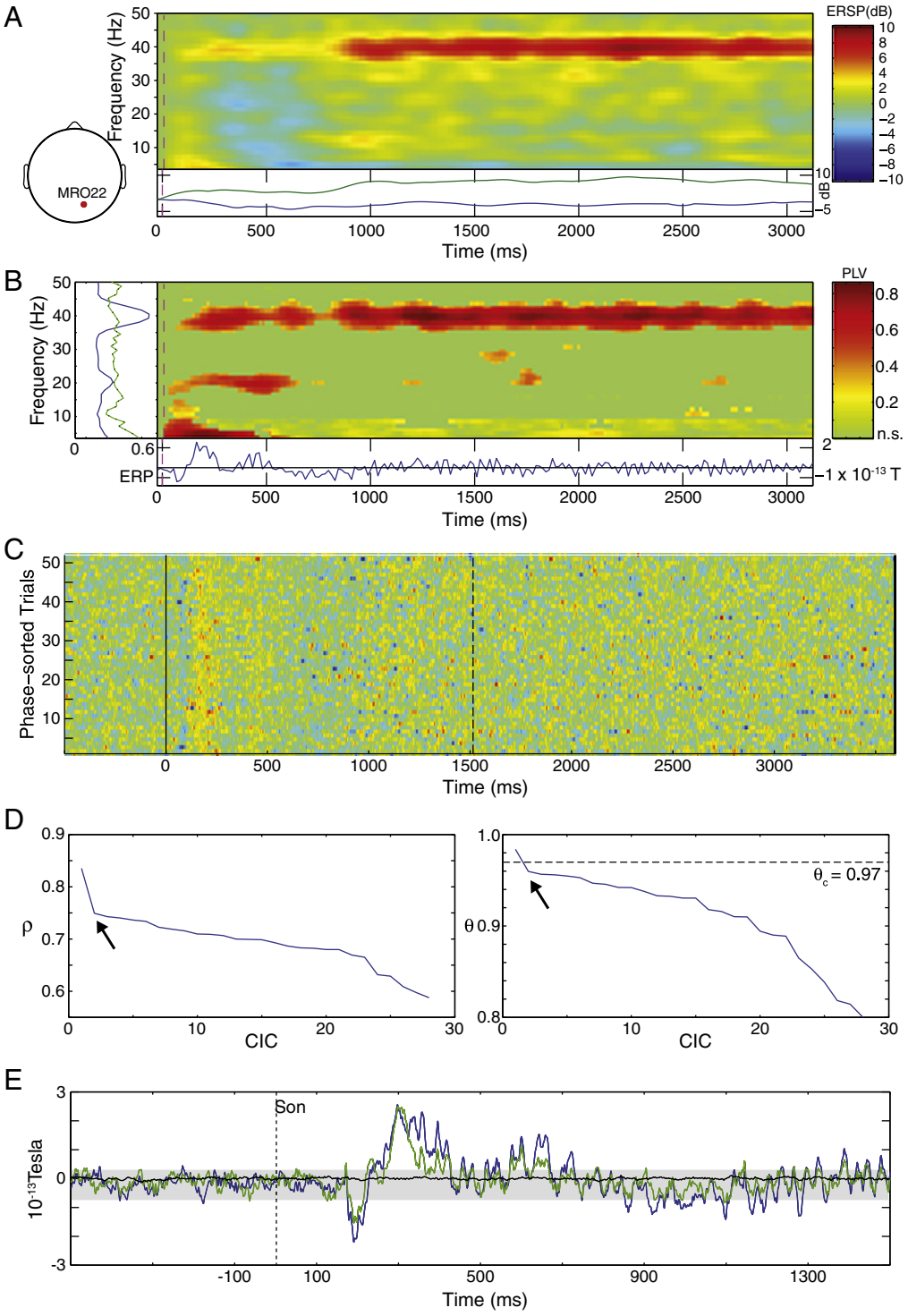


Fig. 6. AECs in SSVEFs (sensor MRO22). (A) Relative spectral power (central part). Time course of min. and max. power (bottom part). (B) Inter-trial phase-locking values (PLVs). Temporal mean of phase-locking values per frequency (left part). Time-frequency distribution of significant PLV tested against surrogate data (central part; $p<0.05$; green color indicates non-significant values). Note high PLVs at the reversal frequency (40.2 Hz). ERF in units of 10^{-13} T (bottom part). (C) Display of single-trial amplitude values. Trials sorted by their phase at 1500 ms (dashed line) at 40.2 Hz from top to bottom. Time within trials from left to right. No clear presence of AECs can be seen. (D) Average correlation coefficient ρ between subset CIC estimates and the corresponding μ-CIC estimate versus component number (left part). Note the drop in correlation from CIC 1 to CIC 2. Stability indices of the estimated CICs (right part). The dashed line indicates a critical stability index Θ_c of 0.97 (based on prior simulation studies — see Fig. 4). Note the drop in the stability index from CIC 1 to CIC 2, indicating particularly stable mixing coefficients for CIC 1. (E) Event-related field at sensor MRO22 (blue) and average 'backprojected' CIC 1 (green) and CIC 2 (black). The minimum-to-maximum amplitude range of CIC 1 in the pre-stimulus baseline period (noise in the time course recovery) is indicated by gray shading. Note the large difference in amplitude between the average 'backprojection' of CIC 1 and that of CIC 2, indicating a lack of phase stability in CIC 2.

each of the chosen critical stability index (see Fig. 4, D). In particular, all estimation runs showed a false-positive rate of less than 0.09, irrespective of the chosen subset size, while the true-positive rate was above 0.95 when estimating more than 4 CICs (i.e. more than 1 additional CICs accounting for the noise residuals) using $\Theta_c<0.99$. No false-positives were reported when estimating more than 8 CICs (see Fig. 4, D). In this case, all true-positively identified AECs had a correlation coefficient $\rho>0.95$ to the simulated original AECs (see example in Fig. 5, D).

In performance evaluations with a fixed threshold for the critical stability index, true-positive rates progressively increased with an increasing number of estimated CICs, while the false-positive rates remained low (Figs. 4, B and D). These findings were in line with our expectation about the effect of estimating additional CICs that account for the noise residuals.

Analysis of LFP recordings revealed ERPs triggered by the onset of a visual stimulus in all three investigated recording sites in visual areas V2, V4 and in primary motor area A4 (Fig. 5). Despite the presence of an event-related potential in all three areas, AECs were only identified in visual areas V2 and V4. Importantly, no AECs were identified in primary motor area A4. These findings were in line with our initial hypothesis about the presence or absence of AECs. In addition, we linearly mixed the obtained AECs using the estimated mixing coefficients ('backprojection') and averaged the resulting signals, thereby obtaining the contribution of additive evoked activity to the original ERP. In both cases (V2, V4) where AECs were identified, their post-stimulus amplitude maxima and minima clearly exceeded pre-stimulus baseline amplitude levels. In addition, they explained a large part, but not all, of the variance (power) in the ERP.

Compatible results were obtained in our MEG study (Fig. 6). We observed a steady-state visual evoked field (SSVEF) response that bore a clear spectral signature of the reversal frequency of the stimulus — this is typical for MEG data (Fawcett et al., 2004) but is in slight contrast with typical results from EEG where the fundamental frequencies dominate (Herrmann, 2001; Vialatte et al., 2010). Time-frequency analysis of magnetic fields at a sensor over the right occipital cortex showed a stimulus related peak at the reversal frequency (40.2 Hz). This response was also characterized by a high phase-locking value (PLVs) over trials, indicating either AECs or contributions of phase-reset oscillatory activity to the SSVEF. A comparison of the task-related increases in spectral power and in PLVs suggested some AECs as part of the SSVEF at 40.2 Hz along with phase-resetting. This is because the time courses of PLVs and of task-related power increases at 40.2 Hz clearly differed with the power increase lagging behind the increase in PLV. However, a decision cannot be made based on these measures, because even PLVs dropped below values of 0.4 at certain times in the stimulus interval (e.g. at 750 ms). This is to be expected because an AEC, if present, may dominate phase calculation at points in time when it has high amplitude, but its contribution to phase calculation is negligible at times when it has small amplitude — it gets buried under spontaneous or induced contributions. Application of our algorithm clearly indicated the presence of 1 AEC. The notion of just one AEC was supported by the fact that only one component's mixing coefficients exceeded the critical stability index $\Theta_c=0.97$. In addition, the first CIC that was estimated was particular with regard to the stability index of its mixing coefficients and also because its average correlation coefficient ρ clearly exceeded the extrapolation of the ρ-curve based on the following CICs (see concave curvature of both curves between CICs 1, 2 and 3 in Fig. 6, D). This fact may serve as a post-hoc verification of the chosen decision criterion. The notion that one and only one AEC is present in the data is also supported by the separate average backprojection of CIC 1 and CIC 2: Only CIC 1 had amplitude values that clearly exceeded the pre-stimulus baseline noise levels (green line and gray shading in Fig. 6, E).

Discussion

There has been a long debate whether the event-related potential (ERP) may be partially generated by phase-resetting of oscillatory activity (e.g. Hanslmayr et al., 2007; Makeig et al., 2002; Sauseng et al., 2007; Sayers et al., 1974; Shah et al., 2004). Studies on this topic mostly attempted to prove that phase-reset oscillations would contribute to the ERP. So far it has not been considered to put the more traditional hypothesis of additive evoked activity as a generator of the ERP to a direct test, i.e. this hypothesis has been assumed to be correct wherever attempts to prove phase-resetting or, lately, asymmetric amplitude modulations of oscillations failed. This may be due to the fact that it was historically the default hypothesis. Here, we demonstrated that the hypothesis of additive evoked contributions (AECs) is testable. Our approach is particular in several ways. First, the use of constrained ICA enables the recovery of large and small amplitude contributions alike, thereby identifying AECs even when contributions of other neuronal activity have large amplitudes in each single-trial (Fig. 6, C). Second, the approach does not rely on the extraction of phase and power values — that is necessarily imprecise in the presence of spontaneous activity. Third, it incorporates information from pre-stimulus baseline and post-stimulus task periods simultaneously. Fourth, it makes use of the multivariate information that is contained in the all-to-all relationship between signals from all single-trials, searching for stereotypical repetition that is best assessed at this multivariate level. This kind of multivariate information has to our knowledge not been used before to solve the problem of identifying AECs. Fifth, the proposed algorithm allows the identification of several independent AECs and to recover their time-courses. While the algorithm can test the presence of additive evoked activity with high sensitivity and specificity, the recovery of the AECs' time courses must stay imperfect due to fundamental mathematical limitations. The pre-stimulus baseline amplitude values of an identified AEC – that theoretically should be zero – may serve as a rough estimate of the remaining noise that could not be removed by the algorithm (see Fig. 6, E).

Our algorithm is able to differentiate between phase-reset oscillations and AECs, which may be surprising if one considers that *in the post-stimulus interval* oscillations that were phase-reset with respect to the stimulus are indistinguishable from AECs. AECs will, however, also have stereotypical – at or near zero – amplitude values in the pre-stimulus baseline period of each trial. Phase-reset oscillations, in contrast, will be identical over trials after the stimulus, but by definition differ in the pre-stimulus period. Hence, phase-reset oscillations will be rejected by our algorithm: The fixed trial-related mixing coefficients of an estimated source signal will not allow contributions of phase-reset oscillatory activity in a single-trial to be split into a phase-stable post-stimulus part and a random-phase pre-stimulus part. It is therefore important to include enough pre-stimulus data into the analysis. In addition, the critical variance index Δ_c may serve to further distinguish AECs and phase reset contributions.

Our algorithm determines the estimated common independent components (CICs) as either additive evoked or not — based on a user-chosen threshold for the critical stability index. This raises the question, what amount of jitter is tolerated before the algorithm will reject a CIC as not being additive evoked. Note that this question should relate the absolute amount of temporal jitter to the dominant frequencies of the signals in question rather than the absolute sampling rate, given signals are adequately sampled. This said, there is a systematic dependency between jitter and the stability indices of the estimated CICs: The larger the jitter the smaller the stability index. Hence, increasing jitter will change a CIC from being accepted as additive evoked to being rejected. This correctly reflects that the characteristics of the component move from additive evoked toward induced. The choice of a maximum tolerable deviation from a true temporal identity over trials is made by the user in the form of a

threshold for the stability index. Nevertheless, this choice is not arbitrary and the need for making this choice is not a weakness of the proposed algorithm but rather a weakness related to the historical definition of AECs as being of 'fixed latency and fixed polarity' to the stimulus. This definition fails to give bounds for what would still be acceptable as being of 'fixed latency'. In the future one might therefore consider thinking about additive evoked and (additive) induced contributions as being two extremes of a continuum and use the stability indices of the CICs as a measure of the 'evokedness' of an estimated additive component. Note that despite this problem AECs can be clearly separated from phase-reset oscillations or spontaneous contributions as detailed above. This is important because Sauseng et al. (2007) recently pointed out that both phase-reset oscillatory activity and additive evoked activity may typically generate the ERP together, albeit with variable contributions and argued that future approaches should try to quantify these contributions. Our method is a first step in this direction.

Few attempts have been made to separate AECs from other contributions or to test for their presence. In a recent review Sauseng et al. (2007) tabulated signal characteristics, that taken together, allow to distinguish ERPs generated by phase-reset oscillations from those generated by AECs better than the application of any single criterion. Nevertheless, they pointed out that none of these criteria is decisive in the sense that it cannot yield false-positives in biologically plausible scenarios. This is expected because all of these criteria involve averaging of some metric – such as time-frequency power distributions – over trials. None of these criteria made use of the multivariate information in the all-to-all relationship among the trial signals. Hence, the decisive property of AECs – their stereotypical presence in each trial – was only evaluated very indirectly. To our knowledge our approach is the first to exploit the multivariate information over trials to quantify the contribution of AECs to the ERP — and each single trial.

Our results suggest that the contribution of AECs to the ERP exists — in line with earlier suggestions in the literature (Shah et al., 2004). However, here we demonstrated the presence of additive evoked activity in the ERP for extra-striate visual areas and also for recordings of neural mass-activity with MEG. The existing slight differences between the SSVEF and the backprojection of the recovered AEC, however, also leave room for the co-existence of another SSVEF generating mechanism such as phase-reset contributions (see also the work of Moratti et al., 2007). Our results also suggest that ERPs can be generated without any contribution of additive evoked activity — especially in cortical areas that are remote from stimulus input, such as primary motor area A4 of the macaque in recordings triggered to a visual stimulus. Our results are compatible with suggestions that ERPs in the primary sensory cortices are to a large part generated by AECs. This can be seen by comparing the variance of the backprojected AECs with the variance of the original ERP in V2 of the macaque (Fig. 5) and also by comparing the backprojected single AEC of the steady-state response with the original VSSEF (Fig. 6, E). Separating contributions of additive evoked activity from contributions of phase-reset oscillatory activity is important to improve our understanding of cortical oscillations and their interplay with the ERP. Ultimately this should help to understand the role of cortical oscillations in top-down modulation of neuronal processing and of cognitive processes as pointed out in Sauseng et al. (2007).

Conclusion

Our algorithm first allows testing for the presence of AECs in each trial. Second, it identifies such contributions even if they are small and only partially generate the ERP. Third, it opens the possibility to decompose observed ERPs into several independent AECs and thus helps to elucidate underlying neuronal mechanisms. This latter possibility may prove useful to understand complex ERP patterns that are thought to arise from the interplay of many neuronal generators.

Acknowledgments

We would like to thank I. Schneider from the MPI for Brain Research, Frankfurt for providing MEG data. We would like to thank T. Ball from the Bernstein Center for Computational Neuroscience, Freiburg for stimulating discussions. MW and GT received support from LOEWE Grant "Neuronale Koordination Foschungsschwerpunkt Frankfurt" (NeFF).

Appendix

The appendix is organized as follows. First, we model a single-trial mixing system for single-channel recorded neuronal activities. We then construct a multichannel mixing system by considering trials as individual sensors, while preserving the additive and stimulus time-locked property of evoked activity on these new sensors. Hereafter we demonstrate that only the additive evoked contributions (AECs) are consistent over diverse sensors subsets and show that the underlying independent source signals of AECs can be recovered by estimating the common independent components (CICs) of the sensor subsets. To this end, we derive an efficient algorithm for estimating CICs and present an approach for testing the reliability of the estimates.

A.1. Single-trial model

In single-trial electrophysiological recordings each trial signal reflects a certain linear mixture of the field potential signals of an unknown number of neuronal current sources. These source signals describe the temporal activity of the underlying neuronal processes, while the mixing process depends on the lead fields which are determined by volume conduction properties of the head (brain, tissue, liquor and bone) and the distances between the sources and the recording electrodes (sensors) (Wolters and de Munck, 2007). Given that we focus on frequencies below 1 kHz and relatively short time courses, we can neglect a convolutive or a time-varying mixing of the source signals during stimulus responses, respectively.

Principle 1. Single-trial mixing process

The mixing process of the underlying source signals from low-frequency transient stimulus responses is instantaneous and stationary on a single-trial level according to the quasistatic approximation of Maxwell's equations.

Let us denote the signal of the j-th trial by the random variable x_j and the i-th source signal that is contributed to x_j by the random variable $\xi_i^{(j)}$. Further on, let M be the number of recorded trials and $N^{(j)}$ the number of source signals contributed to x_j. The single-trial mixing system is then given by

$$x_j = \sum_{i=1}^{N^{(j)}} \alpha_i^{(j)} \cdot \xi_i^{(j)}, \text{ for } \quad j = 1, \ldots, M, \tag{1}$$

where $\alpha_i^{(j)}$ is the mixing coefficient of $\xi_i^{(j)}$ and $\alpha_i^{(j)} \cdot \xi_i^{(j)}$ is referred to as the contribution from the i-th source signal to the j-th trial. Eq. (1) can also be written in vectorized form, if we collect the mixing coefficients in the $N^{(j)} \times 1$ vector $\boldsymbol{\alpha}^{(j)} = (\alpha_1^{(j)}, \ldots, \alpha_{N^{(j)}}^{(j)})^{\mathrm{T}}$ and similarly collect the source signals in the $N^{(j)} \times 1$ vector $\boldsymbol{\xi}^{(j)} = (\xi_1^{(j)}, \ldots, \xi_{N^{(j)}}^{(j)})^{\mathrm{T}}$. The vectorized single-trial mixing system can then be written as

$$x_j = \boldsymbol{\alpha}^{(j)^{\mathrm{T}}} \cdot \boldsymbol{\xi}^{(j)}, \text{ for } \quad j = 1, \ldots, M. \tag{2}$$

In order to distinguish the contributions from source signals that correspond to different types of neural activities, it is reasonable to differentiate between source signals that are contributed to all trials and source signals that are contributed to particular trials only.

Definition 1. Common and individual sources

Let $\xi_i^{(j)}$ be the i-th source signal that is instantaneously and stationary mixed onto the j-th single-trial signal x_j, with $j\in\{1,\ldots,M\}$ and $i\in\{1,\ldots,N^{(j)}\}$. Then, $z_k^{(l)}$ is referred to as the k-th 'common' source signal from the l-th trial if $\forall j\exists!\, i: z_k^{(l)}=\xi_i^{(j)}$ and $y_k^{(l)}$ is referred to as the k-th 'individual' source signal from the l-th trial if $\exists!\,(i,j): y_k^{(l)}=\xi_i^{(j)}:\forall(i',j'):\xi_i^{(j)}\neq\xi_{i'}^{(j')}, j'\in\{1,\ldots,M\}\backslash j,\ i'\in\{1,\ldots,N^{(j')}\}$.

According to Definition 1, each source signal $\xi_i^{(j)}$ can be defined either as a common or an individual source signal. In order to determine the type of a source signal, it is sufficient to simply test whether it is contributed to every trial. This can be done by including all the source signals $\xi_i^{(j)}$ from all trials into the single-trial mixing system and subsequently identifying repeatedly appearing source signals. Such source signals – if present – will reflect common source signals.

Property 1. Reducibility of source signals to unique common and individual source signals

Let $\boldsymbol{\xi}=(\boldsymbol{\xi}^{(1)^{\mathrm{T}}},\ldots,\boldsymbol{\xi}^{(M)^{\mathrm{T}}})^{\mathrm{T}}=(\xi_1^{(1)},\ldots,\xi_{N^{(M)}}^{(M)})^{\mathrm{T}}$ denote the concatenation of all $N=\sum_{j=1}^{M}N^{(j)}$ source signals from M trials. Let further on $\mathbf{z}^{(j)}=(z_1^{(j)},\ldots,z_\Lambda^{(j)})^{\mathrm{T}}$ denote Λ common source signals of $\boldsymbol{\xi}^{(j)}$ and $\mathbf{y}^{(j)}=(y_1^{(j)},\ldots,y_{\Omega^{(j)}}^{(j)})^{\mathrm{T}}$ denote $\Omega^{(j)}$ individual source signals of $\boldsymbol{\xi}^{(j)}$, with $j\in\{1,\ldots,M\}$ and $N^{(j)}=\Lambda+\Omega^{(j)}$. Then, because $z^{(1)}=\ldots=z^{(M)}=z$ it follows that $\mathbf{s}=(\mathbf{z}^{\mathrm{T}},\mathbf{y}^{(1)^{\mathrm{T}}},\ldots,\mathbf{y}^{(M)^{\mathrm{T}}})^{\mathrm{T}}=(s_1,\ldots,s_L)^{\mathrm{T}}$ reflects $L=\Lambda+\sum_{j=1}^{M}\Omega^{(j)}$ 'unique' (i.e. distinct) source signals s_k, with $k=1,\ldots,L$, such that $\forall(i,j)\exists!\,k: s_k=\xi_i^{(j)}$. Consequently, $\mathbf{s}=\mathbf{Q}\cdot\boldsymbol{\xi}$ where $\mathbf{Q}$ is a $L\times N$ row selection matrix with one entry 1 in each row and 0's elsewhere.

Given Property 1, we can now equivalently describe the *j*-th trial signal x_j as a sparse mixture of all the 'unique' source signals. We thus substitute $\boldsymbol{\xi}^{(j)}$ from Eq. (2) by the $L\times 1$ vector $\mathbf{s}=(s_1,\ldots,s_L)^{\mathrm{T}}$ holding all Λ 'common' and $\Omega=\sum_{j=1}^{M}\Omega^{(j)}$ 'individual' source signals, where $L=\Lambda+\Omega$ with $L\leq N$. Similarly, we substitute $\boldsymbol{\alpha}^{(j)}$ by the $L\times 1$ vector $\mathbf{a}^{(j)}=(a_1^{(j)},\ldots,a_L^{(j)})^{\mathrm{T}}$ with sparse entries defined as

$$a_k^{(j)}=\begin{cases}\alpha_i^{(j)} & \text{if } s_k=\xi_i^{(j)}\\ 0 & \text{else}\end{cases},\ \text{with}\quad i\in\left\{1,\ldots,N^{(j)}\right\}. \tag{3}$$

The generalized single-trial mixing system that describes the *j*-th trial signal x_j with respect to all source signals from all trials is then given by

$$x_j=\mathbf{a}^{(j)^{\mathrm{T}}}\cdot\mathbf{s},\ \text{for}\quad j=1,\ldots,M. \tag{4}$$

A.2. Multichannel model

In contrast to the mixing process on the single-trial level, the mixing process on the multitrial level is influenced by trial-to-trial modulations of the sources (e.g. neuronal interactions, plastic changes) as well as small relative movements of cortex and electrodes (e.g. due to blood flow). These movements and modulations cause different contributions from each source signal to the trials and thus a time-varying mixing of the source signals over the trials. That is: The mixing coefficients of each source signal will vary over the trials. In addition, modulations may cause a varying time-delay of each contribution over the trials as well and thus a non-instantaneous mixing. However, according to Definition 1, we can neglect a non-instantaneous mixing, if we consider all time-delayed contributions as contributions from *different individual* source signals.

Property 2. Equality of non-instantaneous and sparse mixtures

Let $\xi[t]$, with $t\in\{1,\ldots,T\}$, be the t-th sample of an arbitrary source signal from an arbitrary trial defined as above, where T is the number of samples. Let further on $\mathscr{T}(\xi,\tau)=\xi[t+\tau]$ denote a time-delay of ξ by τ samples. Then, M non-instantaneous mixtures of $\tilde{\Lambda}$ source signals ξ_i, with $i\in\left\{1,\ldots,\tilde{\Lambda}\right\}$, can be written in vectorized form as $\tilde{\mathbf{x}}=\sum_{i=1}^{\tilde{\Lambda}}\sum_{j=1}^{M}\mathbf{e}_j\cdot\alpha_i^{(j)}\cdot\mathscr{T}\left(\xi_i,\tau_i^{(j)}\right)$, where $\alpha_i^{(j)}$ is the j-th mixing coefficient and $\tau_i^{(j)}$ is the j-th delay factor of ξ_i, respectively, and $\mathbf{e}_j=(e_1,\ldots,e_M)^{\mathrm{T}}$ is the canonical unit vector with one entry 1 at e_j and 0's elsewhere. Thus, by substituting $\mathbf{a}_k=\alpha_i^{(j)}\cdot\mathbf{e}_j$ and $y_k=\mathscr{T}\left(\xi_i,\tau_i^{(j)}\right)$, where $k\rightarrow(i,j)$ reflects a mapping of k onto a unique pair (i,j), with $k\in\left\{1,\ldots,\tilde{\Omega}\right\}$ and $\tilde{\Omega}=M\cdot\tilde{\Lambda}\leq\Omega$, $\tilde{\mathbf{x}}$ can equivalently be written as M instantaneous sparse mixtures of $\tilde{\Omega}$ different individual sources signals given by $\tilde{\mathbf{x}}=\sum_{k=1}^{\tilde{\Omega}}\mathbf{a}_k\cdot y_k$.

Given Property 2, all *M* contributions from $\tilde{\Lambda}$ *time-delayed* source signals can be treated as contributions from $\tilde{\Omega}=M\cdot\tilde{\Lambda}\leq\Omega$ *different individual* source signals. Furthermore, we can neglect a time-varying mixing as well, if we consider the single-trial signals to be simultaneously recorded. This is because, according to Principle 1, such a simultaneous recording describes concurrent single-trial mixing processes that are all stationary. Therefore source signals in a concurrent single-trial mixing system can be considered being mixed with a fixed time-delay and fixed coefficients onto the trials. Consequently, we can construct a virtual multichannel mixing system by treating the trial signals of a single-channel multitrial recording as sensor signals of a simultaneous multichannel recording. To this end, we will refer to the *j*-th trial (signal) as the *j*-th sensor (signal) throughout the rest of the appendix, if not stated otherwise.

Principle 2. Multichannel mixing process

The mixing process of both common and individual source signals is instantaneous and stationary on a multichannel level that is constructed by considering the trials as sensors.

Let us consider the trials to have an equal length *T* and collect the sensor signals in the $M\times 1$ vector $\mathbf{x}=(x_1,\ldots,x_M)^{\mathrm{T}}$ and the corresponding mixing vectors of the *L* source signals in the $M\times L$ matrix $\mathbf{A}=\left[\mathbf{a}^{(1)},\ldots,\mathbf{a}^{(M)}\right]^{\mathrm{T}}$, such that $\mathbf{A}=\left[a_{j,i}\right]$, for $j=1,\ldots,M$ and $i=1,\ldots,L$. The multichannel mixing system written in matrix form is then given by

$$\mathbf{x}=\mathbf{A}\cdot\mathbf{s}. \tag{5}$$

According to Principle 2, Eq. (5) describes concurrent stationary mixing processes allowing to collect all *T* samples of the sensor signals in a $M\times T$ matrix $\mathbf{X}=[\mathbf{x}[1],\ldots,\mathbf{x}[T]]$ and similarly all *T* samples of the source signals in a $L\times T$ matrix $\mathbf{S}=[\mathbf{s}[1],\ldots,\mathbf{s}[T]]$. The time-invariant multichannel mixing system can then be written as

$$\mathbf{X}=\mathbf{A}\cdot\mathbf{S}. \tag{6}$$

Next, let us investigate the mixing of common and individual source signals on the multichannel level. According to Definition 1, common source signals are consistently contributed to all trials and thus, with Principle 2, are *densely* mixed onto the sensors. Individual source signals – on the other hand – are contributed to particular trials only, so they are *sparsely* mixed onto the sensors. Note, that with Property 2, time-delayed contributions may be described by contributions from different individual source signals and thus are considered as sparse mixtures as well. Hence, the multichannel mixing system can equivalently be described as a composed mixing system

consisting of a dense mixing part with regard to Λ common source signals and a sparse mixing part with regard to Ω individual source signals. To this end, we substitute $\mathbf{A} = [\mathbf{B},\mathbf{C}]\cdot\mathbf{P}$ and $\mathbf{S} = \mathbf{P}^{\mathrm{T}}\cdot[\mathbf{Z},\mathbf{Y}]^{\mathrm{T}}$, where $\mathbf{P}$ is a $L\times L$ permutation matrix that sorts the common source signals to obtain the first indices, such that $\mathbf{P}\cdot\mathbf{P}^{\mathrm{T}} = \mathbf{I}$. The composed multichannel mixing system is then given by

$$\mathbf{X} = [\mathbf{B},\mathbf{C}]\cdot\begin{bmatrix}\mathbf{Z}\\ \mathbf{Y}\end{bmatrix} = \mathbf{B}\cdot\mathbf{Z} + \mathbf{C}\cdot\mathbf{Y}, \tag{7}$$

where $\mathbf{Z} = [\mathbf{z}[1],\ldots,\mathbf{z}[T]]$ is a $\Lambda\times T$ row submatrix of $\mathbf{S}$ holding the common source signals z_i, with $i=1,\ldots,\Lambda$, and $\mathbf{Y} = [\mathbf{y}[1],\ldots,\mathbf{y}[T]]$ is the remaining $\Omega\times T$ row submatrix of $\mathbf{S}$ holding the individual source signals y_k, with $k=1,\ldots,\Omega$. Accordingly, $\mathbf{B} = \left[b_{j,i}\right]$ is a dense $M\times\Lambda$ column submatrix of $\mathbf{A}$ holding the mixing coefficients of the common source signals and $\mathbf{C} = \left[c_{j,i}\right]$ is the remaining sparse $M\times\Omega$ column submatrix of $\mathbf{A}$ holding the mixing coefficients of the individual source signals.

Given that $\mathbf{C}$ is sparse each of its columns has exactly one nonzero entry $c^{*}_{j,i}$, while the number of such nonzero entries in each of its rows is given by the number of individual source signals contributing to the corresponding sensor, i.e. $\forall j\exists^{=\Omega^{(j)}} i : c^{*}_{j,i} = \sum_{l=1}^{M} c_{l,i}\neq 0$, with $i\in\{1,\ldots,\Omega\}$ and $j\in\{1,\ldots,M\}$. Hence, we can average columns of $\mathbf{C}$ with nonzero entries in the same row and merge the corresponding individual source signals to 'nuisance' signals without changing the resulting contributions to the sensors at all. This can be done by modifying $\mathbf{C}$ into a $M\times M$ diagonal matrix $\mathbf{D}$ with diagonal entries $d_{j,j} = \left(\Omega^{(j)}\right)^{-1}\cdot\sum_{i=1}^{\Omega} c_{j,i}$, for $j=1,\ldots,M$, and by modifying $\mathbf{Y}$ into a $M\times T$ matrix $\bar{\mathbf{Y}} = [\bar{\mathbf{y}}[1],\ldots,\bar{\mathbf{y}}[T]]$ holding the nuisance signals $\bar{y}_j = \sum_{i=1}^{\Omega}\mathrm{sgn}(c_{j,i})\cdot y_i$, for $j=1,\ldots,M$. Clearly, the benefit of this modification is that a total of Ω individual source signals are reduced to M nuisance signals. Additionally, the contributions from these nuisance signals can be treated as noise signals given by $\mathbf{N} = \mathbf{D}\cdot\bar{\mathbf{Y}} = \mathbf{C}\cdot\mathbf{Y}$. The composed mixing system in Eq. (7) can thus equivalently be described by a noise perturbed mixing system, with the noise signals being decoupled and independent from the densely mixed common source signals. The noise perturbed multichannel mixing system is then given by

$$\mathbf{X} = \mathbf{B}\cdot\mathbf{Z} + \mathbf{N}. \tag{8}$$

Now that we have described the properties of common and individual source signals on the multichannel level, we have to relate their contributions to those from different types of neural activities. According to the explanations given in the methods chapter and Eq. (8), contributions from additive evoked activity are reflected by contributions from common source signals, while contributions from all other types of response activities (i.e. induced and spontaneous) as well as from phase-reset oscillatory activity can be described by noise signals. Note that contributions from asymmetrical oscillatory activity can be considered as a special case of induced activity having a one-sided post-stimulus power increase and will thus not be treated separately in the following.

Principle 3. Contributions from neuronal activity

On the multichannel level additive evoked contributions are reflected by contributions from common source signals, while contributions from time-delayed or sparse activities such as induced, phase-reset or spontaneous activity are described by noise signals.

A.3. Multisubset model

The multichannel mixing system constructed in Section A.2 reflects a generative data model that assumes the sensor signals being generated by common source signals and perturbing noise signals. By combining Definition 1 and Principle 2 we see that this generative data model is equally valid for every subset of the sensor signals. Consequently Principle 3 is valid for any sensor subset as well, allowing the construction of multichannel mixing systems using only subsets of the trial signals.

Let ${}_{M}\boldsymbol{\varphi}^{p}_{\lambda} = \{\varphi^{p}_{1},\ldots,\varphi^{p}_{\lambda}\}$ denote the p-th subset of $\lambda<M$ random indices, where each entry $\varphi^{p}_{j}\rightarrow m$ reflects a unique mapping of index j onto a sensor index m, with $m\in\{1,\ldots,M\}$, such that $\varphi^{p}_{j}\neq\varphi^{p}_{l}$ and ${}_{M}\boldsymbol{\varphi}^{p}_{\lambda}\neq{}_{M}\boldsymbol{\varphi}^{p}_{\lambda}$, where $j\neq l$, $p\neq q$ with $j,l\in\{1,\ldots,\lambda\}$, $p,q\in\{1,\ldots,K\}$. That is: Indices are mutually disjoint in single subsets and mutually almost disjoint over different subsets. Let further on $\mathbf{X}^{\{p\}} = \mathbf{X}_{{}_{M}\boldsymbol{\varphi}^{p}_{\lambda}:}$ denote the p-th sensor subset that consists of the sensor signals given by the ${}_{M}\boldsymbol{\varphi}^{p}_{\lambda}$-th components (rows) of $\mathbf{X}$. Similarly, let $x^{\{p\}}_{j} = x_{\varphi^{p}_{j}}$ denote the j-th sensor signal of the p-th sensor subset that is given by the φ^{p}_{j}-th component (row) of $\mathbf{X}$. The multichannel subset mixing system is then analog to Eq. (8) given by

$$\mathbf{X}^{\{p\}} = \mathbf{B}^{\{p\}}\cdot\mathbf{Z} + \mathbf{N}^{\{p\}}, \tag{9}$$

where $\mathbf{B}^{\{p\}}$ and $\mathbf{N}^{\{p\}}$ denote the mixing coefficients and the noise signals of the p-th sensor subset, respectively. Given that ${}_{M}\boldsymbol{\varphi}^{p}_{\lambda}\neq{}_{M}\boldsymbol{\varphi}^{p}_{\lambda}$ it follows for the noise signals that $\mathbf{N}^{\{p\}}\neq\mathbf{N}^{\{q\}}$, where $p,q\in\{1,\ldots,K\}$ with $p\neq q$, i.e. the noise perturbation is specific to individual sensor subsets $\mathbf{X}^{\{p\}}$. In contrast, since the mixing matrix $\mathbf{B}^{\{p\}}$ is dense, all of the common source signals are consistently contributed to $\mathbf{X}^{\{p\}}$ and consequently to all sensor subsets as well. That is: Common source signals are obviously reflected by the 'common signal components' (Flury, 1984; Schott, 1988) of the sensor subsets. An intuitive approach to reveal these common components would thus be to estimate the mutually maximally similar (in terms of a proper distance measure) signal components from multiple sensor subsets. However, common components estimated in this way would still be biased by the noise perturbation of the sensor subsets. Therefore, it is reasonable to investigate the noise perturbation of the group-average of the sensor subsets, i.e. the average over the entire group of sensor subsets.

Let K almost disjoint sensor subsets according to Eq. (9) consist each of λ distinct sensor signals. The group-average of these sensor subsets is then according to the distributivity of matrix multiplication over matrix addition given by

$$\bar{\mathbf{X}} = \bar{\mathbf{B}}\cdot\mathbf{Z} + \bar{\mathbf{N}}, \tag{10}$$

where $\bar{\mathbf{B}}$ and $\bar{\mathbf{N}}$ are the corresponding group-averages of all $\mathbf{B}^{\{p\}}$ and $\mathbf{N}^{\{p\}}$, respectively, for $p\in\{1,\ldots,K\}$. Let further on $d^{\{p\}}_{j,j} = d_{\varphi^{p}_{j},\varphi^{p}_{j}}$ and $\bar{y}^{\{p\}}_{j} = \bar{y}_{\varphi^{p}_{j}}$ be defined similarly as above as the p-th diagonal entry of $\mathbf{D}^{\{p\}}$ and the p-th component of the subset nuisance signals $\bar{\mathbf{Y}}^{\{p\}}$, respectively, such that $\mathbf{N}^{\{p\}} = \mathbf{D}^{\{p\}}\cdot\bar{\mathbf{Y}}^{\{p\}}$. Then, the group-average noise signals $\bar{\mathbf{N}}$ can accordingly be decomposed into a $\lambda\times\lambda$ diagonal matrix $\tilde{\mathbf{D}}$ with diagonal entries $\tilde{d}_{j,j} = K^{-1}\cdot\sum_{p=1}^{K} d^{\{p\}}_{j,j}$, for $j=1,\ldots,\lambda$, and into a $\lambda\times T$ matrix $\tilde{\mathbf{Y}}$ holding averaged subset nuisance signals $\tilde{y}_j = \sum_{p=1}^{K}\bar{y}^{\{p\}}_{j}$, for $j=1,\ldots,\lambda$, such that $\bar{\mathbf{N}} = \tilde{\mathbf{D}}\cdot\tilde{\mathbf{Y}}$. Furthermore, from ${}_{M}\boldsymbol{\varphi}^{p}_{\lambda}\neq{}_{M}\boldsymbol{\varphi}^{p}_{\lambda}$ it follows that $\forall j\exists(p,q) : \bar{y}^{\{p\}}_{j}\neq\bar{y}^{\{q\}}_{j}$ so that $\tilde{\mathbf{Y}}\neq\bar{\mathbf{Y}}^{\{p\}}$ and thus $\bar{\mathbf{N}}\neq\mathbf{N}^{\{p\}}$, where $p,q\in\{1,\ldots,K\}$ with $p\neq q$.

Property 3. Variance reduction of averaged noise

Let the noise signal $n^{\{p\}}_{j} = d^{\{p\}}_{j,j}\cdot\bar{y}^{\{p\}}_{j}$ and the group-average noise signal $\bar{n}_j = K^{-1}\cdot\sum_{p=1}^{q} d^{\{p\}}_{j,j}\cdot\bar{y}^{\{p\}}_{j}$ be defined as above, for $j\in\{1,\ldots,\lambda\}$. Let further $n^{\{p\}}_{j}$ be mean-free, i.e. $E\{n^{\{p\}}_{j}\} = 0$, *such that* $E\left\{\left(n^{\{p\}}_{j}\right)^{2}\right\} = d^{\{p\}^{2}}_{j,j}\cdot E\left\{\left(\bar{y}^{\{p\}}_{j}\right)^{2}\right\}$ *and* $E\left\{\left(\bar{n}_j\right)^{2}\right\} = E\left\{d^{\{p\}}_{j,j}\cdot d^{\{q\}}_{j,j}\cdot E\left\{\left(\bar{y}^{\{p\}}_{j}\right)\cdot\left(\bar{y}^{\{q\}}_{j}\right)\right\}\right\}$ *reflect the variance of $n^{\{p\}}_{j}$ and $\bar{n}_j$, respectively. Then, since $\forall j\exists(p,q)$:*

$E\left\{\left(\bar{y}_j^{\{p\}}\right)\cdot\left(\bar{y}_j^{\{q\}}\right)\right\} < E\left\{\left(\bar{y}_j^{\{p\}}\right)^2\right\}$ *it follows for the variances that* $E\left\{\left(\bar{n}_j\right)^2\right\} < E\left\{\left(n_j^{\{p\}}\right)^2\right\}$ *and thus* $E\left\{\left(\bar{n}_j\right)\cdot\left(n_j^{\{p\}}\right)\right\}\cdot E\left\{\left(\bar{n}_j\right)^2\right\}^{-1}\cdot E\left\{\left(n_j^{\{p\}}\right)^2\right\}^{-1} < 1$.

Clearly, Property 3 shows that averaging almost disjoint sensor subsets reduces the overall noise variance of the resulting group-average. It should be noted, however, that for an effective noise reduction, the chosen sensor indices in any of the used index subsets have to be arranged in such a way that the signals of the group-average reflect linear combinations of (at least some of) the included sensor signals and are still mutually linearly independent.

A.4. Demixing model

In order to identify the underlying source signals of additive evoked contributions (AECs) we can make use of Principle 3 and attempt recovering the common source signals. Given, however, the conserved consistency of common source signals over all sensor subsets explained in Section A.3, we can equally well attempt recovering the common signal components of the sensor subsets. Our goal in the following will thus be to derive a model for estimating noise reduced common signal components from multiple noise perturbed sensor subsets. (The problem of determining the number of common signal components will be treated in Section A.6.)

Let us first consider the multichannel mixing system given in Eq. (6) and assume that only the Λ common source signals, with $\Lambda = L \leq M$, are contributed to the sensors, i.e. $\mathbf{S} = \mathbf{Z}$ and $\mathbf{A} = \mathbf{B}$. Then, from basic linear algebra it follows that the source signals can be recovered by inverting the mixing process using a demixing matrix $\mathbf{W}$, such that

$$\hat{\mathbf{S}} = \mathbf{W}\cdot\mathbf{X}, \tag{11}$$

where $\hat{\mathbf{S}}$ is a $\lambda\times T$ matrix of the recovered source signals and $\mathbf{W} = \hat{\mathbf{A}}^{+}$ is a $\lambda\times M$ linear estimator for the inverse or pseudo-inverse (if $L<M$) of $\mathbf{B}$, with $\lambda = L$ the number of recovered source signals. Usually $\mathbf{W}$ is unknown but can still be estimated given available *a priori* knowledge of the underlying source signals. In our case, however, we do not have such *a priori* knowledge and thus have to *blindly* estimate a demixing matrix given only the sensor signals (see e.g. Comon and Jutten, 2010). To this end we will make some mild assumptions about the properties of the common source signals that we can use as estimation criteria.

Assumption 1. Statistical independence of common source signals

Let $\mathbf{z} = (z_1, \ldots, z_\Lambda)^{\mathrm{T}}$ *be a vector of* Λ *common source signals defined as above. Then all source signals* z_i, *for* $i = 1,\ldots,\Lambda$, *are assumed non-Gaussian, independently distributed with zero mean and unit variance,* i.e. $E\{z_i\} = 0$, $E\{z_i^2\} = 1$ *and* $p(\mathbf{z}) = \prod_{i=1}^{L} p_i(z_i)$, *where* $p(\mathbf{z}) = p(z_1,\ldots,z_\Lambda)$ *is the joint probability density function (pdf) of all source signals* $z_1,\ldots,z_\Lambda$ *and* $p_i(z_i) = \sum_{z_1,\ldots,z_{i-1},z_{i+1},\ldots,z_L} p(\mathbf{z})$ *is the marginal pdf of the source signal* z_i.

Assumption 1 simply states that we assume that all common source signals are statistically independent from each other. We note that other applications of ICA to neuroscience data yield biologically plausible results, even when compared to other methods of signal decomposition (Wibral et al., 2008), hence this assumption seems to be a fruitful approach to assume some degree of independence.

Given Assumption 1, we can attempt recovering the common source signals with independent component analysis (ICA) (Comon, 1994). In standard ICA the M sensor signals are assumed to be instantaneous linear mixtures of $L \leq M$ hidden source signals that are non-Gaussian and independent identically distributed. The objective of ICA is thus to decompose the sensor signals into λ mutually maximally independent signal components (ICs), where $\lambda \leq M$ is the number assumed sources. Such a decomposition can be achieved by estimating a $\lambda\times M$ demixing matrix $\mathbf{W}_{ICA}$ minimizing the Kullback–Leibler divergence (KLD) between the joint probability density function (pdf) and the product of the marginal pdfs of the estimated signal components given by

$$\begin{aligned}\mathbf{W}_{ICA} &= \arg\min_{\{\mathbf{W}\}}\left(D_{\mathrm{KL}}\left(p\|\prod_j p_j\right)\right)\\ &= \arg\min_{\{\mathbf{W}\}}\left(\sum_{\{\hat{s}_1\},\ldots,\{\hat{s}_M\}} p(\hat{\mathbf{s}})\cdot\log\left(\frac{p(\hat{\mathbf{s}})}{\prod_{j=1}^{M} p_j\left(\hat{s}_j\right)}\right)\right),\end{aligned} \tag{12}$$

where $\hat{s}_j = \mathbf{v}_j^{\mathrm{T}}\cdot\mathbf{x}$ is the j-th IC, with $\mathbf{v}_j$ the j-th column of $\mathbf{W}_{ICA}^{\mathrm{T}}$; $p(\hat{\mathbf{s}})$ is the joint pdf of $\hat{\mathbf{s}}$ and $p_j\left(\hat{s}_j\right)$ is the marginal pdf of $\hat{s}_j$. Since $D_{\mathrm{KL}}\left(p\|\prod_j p_j\right)$ is strictly nonnegative it will attain its minimum if and only if the estimated ICs are mutually statistically independent, i.e. if $p(\hat{\mathbf{s}}) = \prod_{j=1}^{M} p_j\left(\hat{s}_j\right)$. For an in-depth treatment of estimating $\mathbf{W}_{ICA}$ see Hyvärinen et al. (2001).

Next, let us keep in mind that the multichannel mixing system in Eq. (8) is in fact noise perturbed. Decomposing the perturbed sensor signals $\mathbf{X} = \mathbf{B}\cdot\mathbf{Z} + \mathbf{N}$ using ICA would then indeed yield $\lambda = M$ maximally mutually independent signal components, however, these ICs would still reflect M noise biased estimates of Λ common source signals as can be seen from

$$\begin{aligned}\hat{\mathbf{S}}_{ICA} &= \mathbf{W}_{ICA}\cdot[\mathbf{B}\cdot\mathbf{Z} + \mathbf{N}]\\ &= \mathbf{W}_{ICA}\cdot\tilde{\mathbf{X}} + \mathbf{W}_{ICA}\cdot\mathbf{N}\\ &= \hat{\mathbf{Z}}_{ICA} + \hat{\mathbf{N}}_{ICA}\ ,\end{aligned} \tag{13}$$

where $\hat{\mathbf{S}}_{ICA}$ is the $M\times T$ matrix of the estimated ICs, $\tilde{\mathbf{X}} = \mathbf{B}\cdot\mathbf{Z}$ is the $M\times T$ matrix of the noise-free sensor signals; $\hat{\mathbf{Z}}_{ICA}$ is the $M\times T$ matrix of the distorted common source signals and $\hat{\mathbf{N}}_{ICA}$ is the $M\times T$ matrix of the residual noise signals whose variance reflects the bias of the estimation. Note that since both the source recovery and the noise reduction depend on the same demixing matrix $\mathbf{W}_{ICA}$, any IC in Eq. (13) would still reflect a noise biased estimate of a common source signal even if $\mathbf{W}_{ICA}$ were the (pseudo)inverse of $\mathbf{B}$. A noise unbiased estimation can thus only be obtained if *a priori* knowledge about the noise signals is available (see e.g. Hyvärinen, 1999b). Nevertheless, we can exploit the specificity of the noise signals to individual sensor subsets in order to estimate noise reduced common signal components.

For this purpose, let us model the mixing system with respect to the common source signals to be overdetermined, i.e. we assume that $\Lambda < M$, and subdivide the sensor signals into K sensor subsets $\mathbf{X}^{\{p\}} = \mathbf{B}^{\{p\}}\cdot\mathbf{Z} + \mathbf{N}^{\{p\}}$, of size λ each, with $p = 1,\ldots,K$ and $\Lambda \leq \lambda < M$. Then, decomposing each $\mathbf{X}^{\{p\}}$ yield – in analogy to Eq. (13) – corresponding IC subsets $\hat{\mathbf{S}}_{ICA}^{\{p\}}$, each reflecting λ distinctly noise biased estimates of Λ common source signals given by

$$\begin{aligned}\hat{\mathbf{S}}_{ICA}^{\{p\}} &= \mathbf{W}_{ICA}^{\{p\}}\cdot\left[\mathbf{B}^{\{p\}}\cdot\mathbf{Z} + \mathbf{N}^{\{p\}}\right]\\ &= \mathbf{W}_{ICA}^{\{p\}}\cdot\tilde{\mathbf{X}}^{\{p\}} + \mathbf{W}_{ICA}^{\{p\}}\cdot\mathbf{N}^{\{p\}}\\ &= \hat{\mathbf{Z}}_{ICA}^{\{p\}} + \hat{\mathbf{N}}_{ICA}^{\{p\}}\ ,\end{aligned} \tag{14}$$

where – according to the subset notation defined in Section A.3 — the superscript $\{p\}$ indicates the p-th subset of the corresponding matrix. In order to treat the noise bias of this estimation, we will thus make some mild assumptions about the properties of the noise signals as well.

Assumption 2. Cross-correlation and independence of noise signals

Let $\mathbf{n} = (n_1,\ldots,n_M)^{\mathrm{T}}$ *be a vector of* M *noise signals and* $\mathbf{z} = (z_1,\ldots,z_\Lambda)^{\mathrm{T}}$ *be a vector of* Λ *common source signals, both defined as*

above. Then all noise signals n_j, for $j=1,\ldots,M$, are assumed to be mean-free and having non-zero variances σ_j, i.e. *$E\{n_j\}=0$ and $E\{n_j^2\}=\sigma_j>0$. Furthermore, all noise signals are assumed to have non-zero linear cross-correlations and to be statistically independent of the common source signals,* i.e. *$E\{n_j\cdot n_l\}\neq 0$ and $D_{KL}(p_i||q_j)=0$, for $j,l=1,\ldots,M$ and $i=1,\ldots,\Lambda$, where p_i and q_j are the pdf's of z_i and n_j, respectively.*

Assumption 2 states that noise signals are not independent from each other but statistically independent from the common source signals. This seems to be a reasonable assumption because the noise signals and common source signals have very different neuronal origins and can therefore be always made independent.

Applying Assumptions 1 and 2 to Eq. (14), we now see that the common source signals are actually reflected by the common *independent* components (CICs) of the sensor subsets. Our new goal is thus to derive a demixing model for signal components that simultaneously satisfy two criteria: 1) Mutual maximal independence in each sensor subset and 2) Mutual maximal similarity (in a least squares sense) over all sensor subsets. In order to satisfy the latter criterion we can minimize the sum of squared Euclidean distances between the signals of all component subsets $\hat{\mathbf{S}}^{\{p\}}$ and $\hat{\mathbf{S}}^{\{q\}}$, with $p\neq q$, by including the squared Frobenius norm $||\hat{\mathbf{S}}^{\{p\}}-\hat{\mathbf{S}}^{\{q\}}||_F^2$ into Eq. (12). The objective of *common independent component analysis* (CICA) can then be formulated as to estimate K individual $\lambda\times\lambda$ subset demixing matrices $\mathbf{W}_{CICA}^{\{q\}}$ that satisfy

$$\mathbf{W}_{CICA}^{\{q\}}=\underset{\{\mathbf{W}^{\{q\}}\}}{\arg\min}\left(\sum_{\{\hat{s}_1^{\{q\}}\},\ldots,\{\hat{s}_\lambda^{\{q\}}\}}p\left(\hat{\mathbf{s}}^{\{q\}}\right)\cdot\log\left(\frac{p\left(\hat{\mathbf{s}}^{\{q\}}\right)}{\prod_{j=1}^{\lambda}p_j\left(\hat{s}_j^{\{q\}}\right)}\right)+\sum_{\substack{r=1,\\ r\neq q}}^{K}\left\|\hat{\mathbf{S}}^{\{q\}}-\hat{\mathbf{S}}^{\{r\}}\right\|_F^2\right),\tag{15}$$

where $\hat{s}_j^{\{q\}}=\left(\mathbf{v}_j^{\{q\}}\right)^T\cdot\mathbf{x}^{\{q\}}$ is the j-th CIC obtained from the q-th sensor subset, with $\mathbf{v}_j^{\{q\}}$ the j-th column of $\left(\mathbf{W}_{CICA}^{\{q\}}\right)^T$; $p\left(\hat{\mathbf{s}}^{\{q\}}\right)$ is the joint pdf of $\hat{\mathbf{s}}^{\{q\}}$ and $p_j\left(\hat{s}_j^{\{q\}}\right)$ is the marginal pdf of $\hat{s}_j^{\{q\}}$, with $j\in\{1,\ldots,\lambda\}$ and $q\in\{1,\ldots,K\}$. Note, that a necessary constraint when maximizing the similarities of corresponding CICs is to conserve the attained KLD of the component subsets. Hence, we will rewrite Eq. (15) into a form that we can use more efficiently satisfy this constraint.

Let us assume to have K component subsets $\hat{\mathbf{S}}^{\{p\}}=\mathbf{W}^{\{p\}}\cdot\mathbf{X}^{\{p\}}$ estimated using an arbitrary blind estimation method and denote the sum of squared Euclidean distances over all component subsets by $\mathcal{E}=\sum_{p=1}^{K}\sum_{\substack{q=1,\\ q\neq p}}^{K}\left\|\hat{\mathbf{S}}^{\{p\}}-\hat{\mathbf{S}}^{\{q\}}\right\|_F^2$. Then, using the group-average $\tilde{\mathbf{S}}=K^{-1}\cdot\sum_{q=1}^{K}\hat{\mathbf{S}}^{\{q\}}$, $\mathcal{E}$ can equivalently be written as $\mathcal{E}=\sum_{p=1}^{K}\sum_{\substack{q=1,\\ q\neq p}}^{K}\left\|\hat{\mathbf{S}}^{\{p\}}-\tilde{\mathbf{S}}\right\|_F^2=K\cdot\sum_{q=1}^{K}\left\|\hat{\mathbf{S}}^{\{p\}}-\tilde{\mathbf{S}}\right\|_F^2$ if $\hat{\mathbf{S}}^{\{q\}}=\tilde{\mathbf{S}}$ subject to $\mathbf{W}^{\{q\}}\neq\mathbf{0}$, for $q=1,\ldots,K$. Eq. (15) can hence be rewritten to a cost efficient form given by

$$\mathbf{W}_{CICA}^{\{q\}}=\underset{\{\mathbf{W}^{\{q\}}\}}{\arg\min}\left(\sum_{\{\hat{s}_1^{\{q\}}\},\ldots,\{\hat{s}_\lambda^{\{q\}}\}}p\left(\hat{\mathbf{s}}^{\{q\}}\right)\cdot\log\left(\frac{p\left(\hat{\mathbf{s}}^{\{q\}}\right)}{\prod_{j=1}^{\lambda}p_j\left(\hat{s}_j^{\{q\}}\right)}\right)+K\cdot\left\|\hat{\mathbf{S}}^{\{q\}}-\tilde{\mathbf{S}}\right\|_F^2\right).\tag{16}$$

Finally, let us consider the group-average of K CIC subsets $\tilde{\mathbf{S}}_{CICA}=K^{-1}\cdot\sum_{q=1}^{K}\hat{\mathbf{S}}_{CICA}^{\{q\}}$, with $\hat{\mathbf{S}}_{CICA}^{\{p\}}=\mathbf{W}_{CICA}^{\{p\}}\cdot\mathbf{X}^{\{p\}}$, in terms of a noise biased estimate and rewrite $\tilde{\mathbf{S}}_{CICA}$ analog to Eq. (14) as

$$\begin{aligned}\tilde{\mathbf{S}}_{CICA}&=\overline{\mathbf{W}}_{CICA}\cdot\overline{\mathbf{X}}\\&=\overline{\mathbf{W}}_{CICA}\cdot\left(\overline{\mathbf{B}}\cdot\mathbf{Z}\right)+\overline{\mathbf{W}}_{CICA}\cdot\overline{\mathbf{N}}\\&=\tilde{\mathbf{Z}}_{CICA}+\tilde{\mathbf{N}}_{CICA}\ ,\end{aligned}\tag{17}$$

where $\tilde{\mathbf{Z}}_{CICA}$ is the $M\times T$ matrix of the distorted common source signals and $\tilde{\mathbf{N}}_{CICA}$ is the $M\times T$ matrix of the noise signal residuals. Given Property 3 and Assumption 2, it then follows for the variances of the noise signal residuals that $\mathrm{trace}\left(\tilde{\mathbf{N}}_{CICA}\cdot\left(\tilde{\mathbf{N}}_{CICA}\right)^T\right)<\mathrm{trace}\left(\hat{\mathbf{N}}_{CICA}^{\{p\}}\cdot\left(\hat{\mathbf{N}}_{CICA}^{\{p\}}\right)^T\right)$, for $p=1,\ldots,K$, provided that $\mathbf{W}_{CICA}^{\{p\}}$ is scaled in order to conserve the total signal energy, i.e. $\mathrm{diag}\left(\mathbf{W}_{CICA}^{\{p\}}\cdot\left(\mathbf{W}_{CICA}^{\{p\}}\right)^T\right)=\mathbf{I}$. That is: The group-average of the estimated CIC subsets is less noise perturbed than any of the single CIC subsets and consequently reflects a noise reduced estimate of the sought-after common source signals. (The choice of K and λ, influencing the noise reduction and the source identification, respectively, will be treated separately in Section A.6).

A.5. Algorithmic estimation

Given the CICA objective derived in Section A.4, our cardinal task is now to construct an algorithm for estimating the CIC subset demixing matrices $\mathbf{W}_{CICA}^{\{p\}}$, for $p=1,\ldots,K$. To this end, let us consider each $\mathbf{W}_{CICA}^{\{p\}}$ to be invertible and linearly decomposable into two $\lambda\times\lambda$ matrices $\mathbf{V}^{\{p\}}$ and $\mathbf{U}^{\{p\}}$, with $\mathrm{rank}\left(\mathbf{W}_{CICA}^{\{p\}}\right)=\mathrm{rank}\left(\mathbf{U}^{\{p\}}\right)=\mathrm{rank}\left(\mathbf{V}^{\{p\}}\right)=\lambda$, such that $\mathbf{W}_{CICA}^{\{p\}}=\mathbf{U}^{\{p\}}\cdot\mathbf{V}^{\{p\}}$. Then, $\mathbf{W}_{CICA}^{\{p\}}$ can be approximated using a demixing matrix $\mathbf{V}^{\{p\}}$ that minimizes the KL-divergence (KLD) within the p-th component subset, and a regression matrix $\mathbf{U}^{\{p\}}$ that minimizes the residual sum-of-squares (RSS) between the p-th component subset and the group-average, while conserving the KLD attained by $\mathbf{V}^{\{p\}}$. For this purpose, we next construct an alternating least squares (ALS) approach (see e.g. Kiers and ten Berge, 1989) consisting of an iterative scheme of alternated reestimations of all $\mathbf{V}^{\{p\}}$ and all $\mathbf{U}^{\{p\}}$, including the above constraints.

In each ALS iteration we first update $\hat{\mathbf{S}}_{CICA}^{\{p\}}$ using the current $\mathbf{W}_{CICA}^{\{p\}}$ and estimate a new $\mathbf{V}^{\{p\}}$ based on a single KLD reduction step regarding $\hat{\mathbf{S}}_{CICA}^{\{p\}}$, while keeping $\mathbf{U}^{\{p\}}$ and $\hat{\mathbf{S}}_{CICA}^{\{p\}}$ fixed. Hereafter we update $\hat{\mathbf{S}}_{CICA}^{\{p\}}$ using the obtained $\mathbf{V}^{\{p\}}$ and estimate a new $\mathbf{U}^{\{p\}}$ based on a single RSS reduction step regarding $\mathbf{S}_{CICA}$, while keeping $\mathbf{V}^{\{p\}}$ and $\hat{\mathbf{S}}_{CICA}^{\{p\}}$ fixed. Finally we update $\hat{\mathbf{S}}_{CICA}^{\{p\}}$ using the obtained $\mathbf{U}^{\{p\}}$ and compute a new $\mathbf{W}_{CICA}^{\{p\}}$ based on $\mathbf{V}^{\{p\}}$ and $\mathbf{U}^{\{p\}}$. These steps are repeated for all K component subsets whereupon $\tilde{\mathbf{S}}_{CICA}$ is updated by computing the group-average of all obtained $\hat{\mathbf{S}}_{CICA}^{\{p\}}$. The updated $\tilde{\mathbf{S}}_{CICA}$ is then used in the RSS reduction step of the next iteration. In order to satisfy the criterion of uncorrelated signal components in ICA, we decorrelate the signals of each sensor subset prior to all iterations using a corresponding $\lambda\times\lambda$ subset whitening matrix $\mathbf{Q}^{\{p\}}$ (see e.g. Cichocki and Amari, 2002) and during the iterations after each KLD and RSS reduction step using a symmetric orthonormalization (Löwdin, 1950) of $\mathbf{V}^{\{p\}}$ and $\mathbf{U}^{\{p\}}$, respectively. The outlined ALS iteration scheme is then continued until algorithmic convergence is achieved, i.e. no significant change in all of the $\mathbf{W}_{CICA}^{\{p\}}$ is obtained.

Given the fast convergence of fixed point iterations we use for the KLD reduction step a single gradient step of FastICA (Hyvärinen, 1999a). For the RSS reduction step we use a single least-squares regression step of a modified EM-ePCA (Ahn et al., 2007), aimed at minimizing (in a least squares sense) the integrated reconstruction error (IRE) (Choi et al., 2006) of the common orthogonal signal subspace of the p-th component subset and the component group-average

$$\begin{aligned}&\mathbf{U}_{LS}^{\{p\}}=\underset{\{\mathbf{U}^{\{p\}}\}}{\arg\min}\left(\sum_{i=j}^{\lambda}\alpha_j\cdot\left\|\mathbf{U}^{\{p\}}\cdot\mathbf{E}_j\cdot\hat{\mathbf{S}}_{CICA}^{\{p\}}-\tilde{\mathbf{S}}_{CICA}\right\|_F^2\right),\\&\text{subject to}:\quad\mathbf{U}_{LS}^{\{p\}}\cdot\left(\mathbf{U}_{LS}^{\{p\}}\right)^T=\mathbf{I},\quad p=1,\ldots,K,\end{aligned}\tag{18}$$

where $\mathbf{U}_{LS}^{\{p\}}$ is the p-th $\lambda\times\lambda$ least-squares regression matrix, $\mathbf{E}_j$ is the p-th $\lambda\times\lambda$ diagonal matrix with diagonal entries $\left[\mathbf{E}_j\right]_{kk}=1$, if $k\leq j$ and else $\left[\mathbf{E}_j\right]_{kk}=0$, for $j,k=1,\ldots,\lambda$, and $\alpha_j>0$ are positive coefficients. The minimization of the IRE in Eq. (18) essentially describes an asymmetric PCA (APCA) (Diamantaras and Kung, 1996) objective of estimating the exact principal singular subspace (Kaiser et al., 2010) of

the p-th subset cross-covariance matrix $\mathbf{R}^{\{p\}} = \hat{\mathbf{S}}_{CICA}^{\{p\}} \cdot \left(\tilde{\mathbf{S}}_{CICA}\right)^{\mathrm{T}}$. Hence, we modify the M-Step of EM-ePCA with regard to APCA using the propositions for nonsymmetric matrices in Hasan (2006). The RSS reduction step is then given by the M-Step of the resulting EM-eAPCA that – in analogy to the *limiting case* of EM-ePCA – can be derived as

$$\mathbf{U}_{LS}^{\{p\}} \leftarrow \mathbf{U}_{LS}^{\{p\}} \cdot \left[UT\left(\mathbf{R}^{\{p\}} + \left(\mathbf{R}^{\{p\}^{\mathrm{T}}}\right)\right)^{-}\right], \tag{19}$$

where $\mathbf{R}^{\{p\}} + \left(\mathbf{R}^{\{p\}^{\mathrm{T}}}\right)$ is used for reason of symmetry and $UT(.)$ denotes the upper triangular matrix.

Next, let us treat the constraint of conserving the attained KLD of the subset components during the RSS reduction step. In Oja (1997) it was shown that nonlinear PCA of uncorrelated signals is equivalent to ICA if an appropriate nonlinear contrast function is used to maximize the mutual independence. Consequently, nonlinearizing the M-Step of EM-eAPCA in Eq. (19) will yield subset components that have a maximal statistical cross-dependency to corresponding group-average components. Since this criterion is by definition best met for subset CICs, their attained KLD will obviously be conserved as well. To this end, we substitute $\mathbf{R}^{\{p\}}$ in Eq. (19) by the sum of two nonlinear subset cross-covariance matrices $\mathbf{R}_1^{\{p\}} = g\left(\hat{\mathbf{S}}_{CICA}^{\{p\}}\right) \cdot \left(\tilde{\mathbf{S}}_{CICA}\right)^{\mathrm{T}}$ and $\mathbf{R}_2^{\{p\}} = \hat{\mathbf{S}}_{CICA}^{\{p\}} \cdot g\left(\tilde{\mathbf{S}}_{CICA}\right)^{\mathrm{T}}$, where $g(.)$ is chosen similarly to Karhunen and Ukkonen (2007) as the *tanh* contrast function. The nonlinear RSS (nRSS) reduction step is then given by

$$\mathbf{U}_{LS}^{\{p\}} \leftarrow \mathbf{U}_{LS}^{\{p\}} \cdot \left[UT\left(\left(\mathbf{R}_1^{\{p\}} + \mathbf{R}_2^{\{p\}}\right) + \left(\mathbf{R}_1^{\{p\}} + \mathbf{R}_2^{\{p\}}\right)^{\mathrm{T}}\right)^{-}\right], \tag{20}$$

where again $\left(\mathbf{R}_1^{\{p\}} + \mathbf{R}_2^{\{p\}}\right) + \left(\mathbf{R}_1^{\{p\}} + \mathbf{R}_2^{\{p\}}\right)^{\mathrm{T}}$ is used for the reason of symmetry.

Finally let us consider two crucial algorithmic aspects: The estimation of a global optimum and the stability of the algorithmic convergence. For the first, recall the validity of the generative data model for any sensor subset explained in Section A.3. Then, from Eq. (9) we see that each of the Λ mixing coefficients from all rows of the mixing matrices $\mathbf{B}^{\{p\}}$, $p=1,\ldots,K$, that correspond to the same sensor indices has a unique value over all $\mathbf{B}^{\{p\}}$. These mixing coefficients reflect a particular component of the basis vectors spanning the *common sources subspace*. Consequently, when estimating $\lambda \geq \Lambda$ CICs, all $\lambda \times \lambda$ CIC subset mixing matrices $\mathbf{A}_{CICA}^{\{p\}} = \left(\mathbf{W}_{CICA}^{\{p\}}\right)^{-}$, $p=1,\ldots,K$, have to share similar entries (estimated mixing coefficients) among each other in those rows that correspond to the same sensor indices. The similarity of such multiply estimated mixing coefficients, however, is influenced by the distinct variance levels of the noise perturbations over the sensor subsets. We maximize the similarity of multiply estimated mixing coefficients by substituting rows corresponding to the same sensor indices over all $\mathbf{A}_{CICA}^{\{p\}}$ by an *a posteriori* estimate reflecting the respective group-average of the mixing coefficients. According to our objective of estimating the group-average of the CICs (see Section A.4), the *a posteriori* estimate can be chosen as the mean over the respective rows.

The stability of the convergence can further be enhanced by decoupling the CIC group-average $\tilde{\mathbf{S}}_{CICA}$ in the nRSS reduction step from the subset matrices $\mathbf{W}_{CICA}^{\{p\}}$, $\mathbf{Q}^{\{p\}}$ and $\mathbf{X}^{\{p\}}$. To this end, we substitute $\tilde{\mathbf{S}}_{CICA}$ in the nRSS reduction step by the p-excluded CIC group-average $\tilde{\mathbf{S}}(-p)_{CICA} = (K-1)^{-1} \cdot \sum_{q \neq p} \hat{\mathbf{S}}_{CICA}^{\{p\}}$, i.e. by the mean over all but the p-th CIC subset. This way, even if $\tilde{\mathbf{S}}(-p)_{CICA}$ became zero, the nRSS reduction step would still monotonically minimize the objective in Eq. (16) (Ten Berge, 1977). The relationship between the similarity criterion using $\tilde{\mathbf{S}}_{CICA}$ and $\tilde{\mathbf{S}}(-p)_{CICA}$ is given by $\| \tilde{\mathbf{S}}_{CICA} - \hat{\mathbf{S}}_{CICA}^{\{q\}} \|_{\mathrm{F}}^2 = \left(\frac{K-1}{K}\right)^{-1} \| \tilde{\mathbf{S}}(-p)_{CICA} - \hat{\mathbf{S}}_{CICA}^{\{q\}} \|_{\mathrm{F}}^2$.

The CICA algorithm can finally be summarized in pseudocode as follows:

Algorithm CICA:

for $p=1:K$
$\mathbf{Q}^{\{p\}} \leftarrow \left(\mathbf{R}_X^{\{p\}}\right)^{-1/2}$
$\mathbf{W}_{CICA}^{\{p\}} \leftarrow \mathbf{Q}^{\{p\}}$
end
for $p=1:K$
$\hat{\mathbf{S}}_{CICA}^{\{p\}} \mathbf{W}_{CICA}^{\{p\}} \mathbf{X}^{\{p\}}$
$\tilde{\mathbf{S}}(\mathbf{p})_{CICA} \left(\mathbf{K} \tilde{\mathbf{S}}_{CICA} \hat{\mathbf{S}}_{CICA}^{\{p\}}\right) / (\mathbf{K}1)$
$\mathbf{V}^{\{p\}} \leftarrow \arg\min_{\{\mathbf{v}^{(p)}\}} \left(D_{\mathrm{KL}}\left(q^{\{p\}} || \prod_j q_j^{\{p\}}\right)\right)$
$\mathbf{V}^{\{p\}} \leftarrow \left(\mathbf{V}^{\{p\}} \cdot \left(\mathbf{V}^{\{p\}}\right)^{\mathrm{T}}\right)^{-1/2} \cdot \mathbf{V}^{\{p\}}$
$\hat{\mathbf{S}}_{ICA}^{\{p\}} \mathbf{V}^{\{p\}} \mathbf{Q}^{\{p\}} \mathbf{X}^{\{p\}}$
$\mathbf{U}_{LS}^{\{p\}} \leftarrow \mathbf{U}_{LS}^{\{p\}} \cdot \left[UT\left(\left(\mathbf{R}_1^{\{p\}} + \mathbf{R}_2^{\{p\}}\right) + \left(\mathbf{R}_1^{\{p\}} + \mathbf{R}_2^{\{p\}}\right)^{\mathrm{T}}\right)^{-}\right]$
$\mathbf{U}_{LS}^{\{p\}} \leftarrow \left(\mathbf{U}_{LS}^{\{p\}} \cdot \left(\mathbf{U}_{LS}^{\{p\}}\right)^{\mathrm{T}}\right)^{-1/2} \cdot \mathbf{U}_{LS}^{\{p\}}$
$\mathbf{W}_{CICA}^{\{p\}} \leftarrow \mathbf{U}_{LS}^{\{p\}} \cdot \mathbf{V}^{\{p\}} \cdot \mathbf{Q}^{\{p\}}$
$\tilde{\mathbf{S}}_{CICA} \leftarrow \left((K-1) \cdot \tilde{\mathbf{S}}(-p)_{CICA} + \mathbf{W}_{CICA}^{\{p\}} \cdot \mathbf{X}^{\{p\}}\right) / K$
end
$\mathbf{A} \leftarrow 0$
for $p=1:K$
$\mathbf{A}_{M\varphi_\lambda^p,:} \leftarrow \mathbf{A}_{M\varphi_\lambda^p,:} + \left(\mathbf{W}_{CICA}^{\{p\}}\right)^{-}$
end
$\mathbf{A} \leftarrow \mathbf{D}(K) \cdot \mathbf{A}$
$\tilde{\mathbf{S}}_{CICA} \leftarrow K^{-1} \cdot \sum_{p=1}^{K} \left(\mathbf{A}_{M\varphi_\lambda^p,:}\right)^{-} \cdot \mathbf{X}^{\{p\}}$

A.6. Test of the algorithmic estimation

In this section we first describe a straightforward method to test for stable CICs, i.e. for indicating the presence of common sources. We then describe the choice of the number and size of the created sensor subsets necessary for the identification of the common source signals.

Given the validity of the generative data model for sensor subsets (see Section A.3) the signal space of every sensor subset has to include a subspace that accommodates the common source signals. When using almost disjoint sensor subsets, the components of the basis vectors spanning this subspace are repeatedly described by sets of entries in those rows of the CIC subset mixing matrices that correspond to the same sensor signals. According to the explanations given in Section A.5, such sets of entries are only stable (i.e. have close to unique values) if they reflect multiple estimates for the mixing coefficients of the common source signals and if the noise perturbations in the sensor subsets is not too high. Hence, stable CIC mixing coefficients indicate the presence of a common sources subspace and consequently the presence of common sources as well. In order to obtain a measurement to test for stable CIC mixing coefficients, we can use the average deviations of subset coefficients from their corresponding group-average coefficients. The distinction between stable and instable CIC mixing coefficients is then a task of determining a proper deviation threshold that discriminates a group of low from a group of high deviations.

Let us now treat the problem of choosing the size of the created sensor subsets. From the explanations given above we know that common source signals are reflected by CICs with stable mixing coefficients,

provided that the generative data model holds and the noise perturbations are not too high. The identification of such CICs, however, depends on the estimation of the common sources subspace. From linear algebra the basis vectors spanning this subspace can only be estimated if the subset size λ is at least as large as the number of the common sources Λ, i.e. $\Lambda \leq \lambda < M$. A subset size for a reliable identification of the common source signals can thus be found by means of determining stable CIC mixing coefficients in repeated analyses with increasing number of estimated CICs. This way the obtained solutions will disclose a transition to solution configurations that consistently comprise a subset of the same stable CICs. The absence of such a transition is then either an indicator that the number of common sources is larger than the number sensor signals, i.e. $\Lambda > M$, or that the generative data model does not hold. While the latter case is indicated by an instability of the CIC mixing coefficients in all analyses, the former case is indicated by a progressively increased stability, based on a subspace approximation of increased degree of freedom.

Finally let us treat the number of the created sensor subsets. In order to accurately estimate the common sources subspace we have to exploit all available multivariate information in the multichannel sensor signals. That is, we have to create sensor subsets that collectively exhaust all available sensor signals. Knowing also that the convergence of CICA is improved when using almost disjoint sensor subsets, the minimum number of sensor subsets is given by $K = \lfloor M/\lambda + 1 \rfloor$, while the maximum number is given by $K = \binom{M}{\lambda}$, with $\lambda < M$. Note, however, that the (theoretical) total number of sensor arrangements allowed under the above constraints is given by $\overline{K} = \left[\binom{M}{\lambda} \cdot \lambda!\right] - 1 \gg K$. A more thorough analysis of subspace approximability and source identifiability on the basis of the size and number of created sensor subsets can be found in the supplementary materials of von Bünau et al. (2009).

B.1. Technical note on the use of the fixed polarity constraint

The algorithm the form presented here only requires a fixed waveform at a fixed latency with respect to the event that lead to a response. That means we use only a weak form of the fixed polarity constraint. In principle our algorithm will also recover sign flipped signal components that otherwise are exactly identical over sensors (i.e. trials). For the intended application of the algorithm to event-related potentials, however, this does not present a problem. For reasons of cellular and network physiology the occurrence of such components in actual neuronal recordings seems extremely improbable, as the underlying physiological mechanisms (ion pumps and channels, excitatory versus inhibitory connectivity patterns, size of EPSPs versus (shunting) IPSPs) seem to be too asymmetric for this situation to occur, especially without additional latency shifts. That is, most likely, a sign flip would change the amplitudes of different parts of a wave shape with different gains going from excitatory to inhibitory mechanisms and vice versa. Nevertheless, we could add an additional constraint to enforce also a strong fixed polarity criterion. Practically, however, this will create problems with convergence and local minima in cases of low signal-to-noise ratio. If it is suspected that sign flipped copies of a signal exist in the mixtures then one could inspect the signs of the corresponding column of the reconstructed (global) mixing matrix.

References

Ahn, J., Oh, J., Choi, S., 2007. Learning principal directions: integrated-squared-error minimization. Neurocomputing 70, 1372–1381.

Bertrand, O., Tallon-Baudry, C., 2000. Oscillatory gamma activity in humans: a possible role for object representation. Int. J. Psychophysiol. 38, 211–223.

Bressler, S.L., Coppola, R., Nakamura, R., 1993. Episodic multiregional cortical coherence at multiple frequencies during visual task performance. Nature 366, 153–156.

Choi, S., Ahn, J.-H., Cichocki, A., 2006. Constrained projection approximation algorithms for principal component analysis. Neural Process. Lett. 24, 53–65.

Cichocki, A., Amari, S., 2002. Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications. John Wiley and Sons, Inc.

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

Comon, P., Jutten, C., 2010. Handbook of Blind Source Separation. Academic Press.

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

Diamantaras, K.I., Kung, S.Y., 1996. Principal Component Neural Networks: Theory and Applications. John Wiley & Sons, Inc.

Fawcett, I.P., Barnes, G.R., Hillebrand, A., Singh, K.D., 2004. The temporal frequency tuning of human visual cortex investigated using synthetic aperture magnetometry. Neuroimage 21, 1542–1553.

Flury, B.N., 1984. Common principal components in K groups. J. Am. Stat. Assoc. 79, 892–898.

Galambos, R., 1992. A comparison of certain gamma band 40 Hz brain rhythms in cat and man. In: Basar, E., Bullock, T.H. (Eds.), Induced Rhythms in the Brain. Birkhauser, Boston, Basel, Berlin, pp. 201–216.

Galambos, R., Sheatz, G.C., 1962. An electroencephalograph study of classical conditioning. Am. J. Physiol. 203, 173–184.

Hanslmayr, S., Klimesch, W., Sauseng, P., Gruber, W., Doppelmayr, M., Freunberger, R., Pecherstorfer, T., Birbaumer, N., 2007. Alpha phase reset contributes to the generation of ERPs. Cereb. Cortex 17, 1–8.

Hasan, M.A., 2006. Sanger's like systems for generalized principal and minor component analysis. Forth IEEE Workshop on Sensor Array and Multichannel Processing, pp. 425–429.

Herrmann, C.S., 2001. Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. Exp. Brain Res. 137, 346–353.

Hyvärinen, A., 1999a. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Netw. 10, 626–634.

Hyvärinen, A., 1999b. Fast ICA for noisy data using Gaussian moments. Proceeding of the 1999 IEEE International Symposium on Circuits and Systems, pp. 57–61.

Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural Netw. 13, 411–430.

Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. Wiley-Interscience.

Kaiser, A., Schenck, W., Moller, R., 2010. Coupled singular value decomposition of a cross-covariance matrix. Int. J. Neural Syst. 20, 293–318.

Karhunen, J., Ukkonen, T., 2007. Extending ICA for finding jointly dependent components from two related data sets. Neurocomputing 70, 2969–2979.

Kiers, H., ten Berge, J., 1989. Alternating least squares algorithms for simultaneous components analysis with equal component weight matrices in two or more populations. Psychometrika 54, 467–473.

Koldovsky, Z., Tichavsky, P., Oja, E., 2006. Efficient variant of algorithm FastICA for independent component analysis attaining the Cramér–Rao lower bound. IEEE Trans. Neural Netw. 17, 1265–1277.

Lewicki, M.S., Sejnowski, T.J., 2000. Learning overcomplete representations. Neural Comput. 12, 337–365.

Löwdin, P., 1950. On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals. J. Chem. Phys. 18, 365–375.

Makeig, S., Westerfield, M., Jung, T.P., Enghoff, S., Townsend, J., Courchesne, E., Sejnowski, T.J., 2002. Dynamic brain sources of visual evoked responses. Science 295, 690–694.

Mazaheri, A., Jensen, O., 2008. Asymmetric amplitude modulations of brain oscillations generate slow evoked responses. J. Neurosci. 28, 7781–7787.

Merchant, H., Battaglia-Mayer, A., Georgopoulos, A.P., 2001. Effects of optic flow in motor cortex and area 7a. J. Neurophysiol. 86, 1937–1954.

Moratti, S., Clementz, B.A., Gao, Y., Ortiz, T., Keil, A., 2007. Neural mechanisms of evoked oscillations: stability and interaction with transient events. Hum. Brain Mapp. 28, 1318–1333.

Nikulin, V.V., Linkenkaer-Hansen, K., Nolte, G., Lemm, S., Muller, K.R., Ilmoniemi, R.J., Curio, G., 2007. A novel mechanism for evoked responses in the human brain. Eur. J. Neurosci. 25, 3146–3154.

Oja, E., 1997. The nonlinear PCA learning rule in independent component analysis. Neurocomputing 17, 25–45.

Regan, D., 1977. Steady-state evoked potentials. J. Opt. Soc. Am. 67, 1475–1489.

Sauseng, P., Klimesch, W., Gruber, W.R., Hanslmayr, S., Freunberger, R., Doppelmayr, M., 2007. Are event-related potential components generated by phase resetting of brain oscillations? A critical discussion. Neuroscience 146, 1435–1444.

Sayers, B.M., Beagley, H.A., Henshall, W.R., 1974. The mechansim of auditory evoked EEG responses. Nature 247, 481–483.

Schott, J.R., 1988. Common principal component subspaces in two groups. Biometrika 75, 229–236.

Shah, A.S., Bressler, S.L., Knuth, K.H., Ding, M., Mehta, A.D., Ulbert, I., Schroeder, C.E., 2004. Neural dynamics and the fundamental mechanisms of event-related brain potentials. Cereb. Cortex 14, 476–483.

Ten Berge, J., 1977. Orthogonal procrustes rotation for two or more matrices. Psychometrika 42, 267–276.

Vialatte, F.B., Maurice, M., Dauwels, J., Cichocki, A., 2010. Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. Prog. Neurobiol. 90, 418–438.

von Bünau, P., Meinecke, F.C., Király, F.C., Müller, K.-R., 2009. Finding stationary subspaces in multivariate time series. Phys. Rev. Lett. 103, 214101.

Wibral, M., Turi, G., Linden, D.E., Kaiser, J., Bledowski, C., 2008. Decomposition of working memory-related scalp ERPs: crossvalidation of fMRI-constrained source analysis and ICA. Int. J. Psychophysiol. 67, 200–211.

Wolters, C., de Munck, J.C., 2007. Volume conduction. Scholarpedia 2, 1738.