Artifact correction in continuous recordings of the electro- and magnetoencephalogram by spatial filtering

Inauguraldissertation

zur Erlangung des akademischen Grades

eines Doktors der Naturwissenschaften

der Universität Mannheim

vorgelegt von

Nicole Ille

aus Göppingen

April 2001

Dekan: Professor Dr. Guido Moerkotte, Universität Mannheim

Referent: Professor Dr. Reinhard Männer, Universität Mannheim

Korreferent: Professor Dr. Michael Scherg, Universität Heidelberg

Tag der mündlichen Prüfung: 09. Juli 2001

Zusammenfassung

In der vorliegenden Dissertation werden zwei neue räumliche Filterverfahren zur Artefaktkorrektur in kontinuierlichen Aufzeichnungen des Elektro- (EEG) und Magnetoenzephalogramms (MEG) dargestellt. Oberstes Ziel der Artefaktkorrektur ist es, Artefakte komplett zu entfernen ohne relevante Hirnaktivität zu verzerren.

Die beiden vorgestellten räumlichen Filter unterscheiden sich von früheren räumlichen Filterverfahren wie dem multiplen Quellenansatz zur (Augen-)Artefaktkorrektur (engl. multiple source eye correction, Abk. MSEC) (Berg und Scherg, 1991a,b, 1994) und der Unabhängigkeitsanalyse (engl. independent component analysis, Abk. ICA) (z.B. Vigário et al., 1998; Jung et al., 1998, 2000a,b) in den Techniken, die zur Modellierung von Artefakt- und Signalraum verwendet werden. Artefakttopographien werden wie in der MSEC-Methode von einzelnen oder gemittelten Artefaktprototypen abgeleitet. Zur Schätzung des Signalraums wurden im Rahmen der Doktorarbeit zwei neue Verfahren entwickelt: der Präselektions-Ansatz (Ille et al., 1997) und die Unabhängigkeitsanalyse mit räumlicher Nebenbedingung (engl. spatially constrained ICA, Abk. SCICA) (Ille et al., 2001).

Im Präselektions-Ansatz werden die Signaltopographien in einer Eigenvektor-Zerlegung aus einem artefaktfreien Ausschnitt des zu korrigierenden Datensegments ermittelt. Um den artefaktfreien Ausschnitt zu erzeugen, werden Signaltopographien aus dem gemessenen Datensegment ausgeschlossen, die einen Amplitudenschwellwert oder eine maximale Korrelation mit dem vordefinierten Artefaktraum überschreiten.

In der SCICA-Methode werden die Signaltopographien aus dem gesamten zu korrigierenden Datensegment geschätzt. SCICA nutzt das Vorwissen über die Artefakttopographien und kombiniert dieses mit der zeitlich-statistischen Vorgehensweise der ICA, um die Signaltopographien zu ermitteln. Ausgehend von n Artefakttopographien wird ein m-kanäliges Datensegment iterativ in m-n weitere Komponenten zerlegt, so dass die Wellenformen aller Komponenten unter der räumlichen Nebenbedingung maximal unabhängig sind. Ein iterativer Algorithmus zur Berechnung der SCICA-Zerlegung wird vorgestellt.

Anwendung von räumlicher Filterung mit Präselektion und SCICA auf EEG-/MEG-Segmente zeigt, dass beide Verfahren in der Lage sind, Artefakte komplett zu entfernen ohne dabei relevante Hirnaktivität zu zerstören. Die Qualität der räumlichen Filterung mit Präselektion hängt jedoch entscheidend von dem gewählten Amplituden- und Korrelationsschwellwert sowie der Anzahl der verwendeten Eigenvektoren ab. Vorteile der SCICA-Zerlegung gegenüber der herkömmlichen ICA werden dargestellt.

Contents

1. Introduction 1

2. Artifacts 5
   - 2.1. Physiological artifacts 5
     - 2.1.1. Ocular artifacts 5
     - 2.1.2. Cardiac artifacts 8
     - 2.1.3. Muscle artifacts 9
     - 2.1.4. Other physiological artifacts 10
   - 2.2. Non-physiological artifacts 11

3. Existing methods for artifact correction 13
   - 3.1. EOG subtraction 13
     - 3.1.1. Time and frequency domain approach 14
     - 3.1.2. Scaling factors and functions 14
     - 3.1.3. Subtracting the EOG 15
   - 3.2. Projection method 15
     - 3.2.1. Projection method for artifact correction 16
     - 3.2.2. Spatial filter 17
   - 3.3. Multiple source eye correction (MSEC) 17
     - 3.3.1. Multiple source approach 18
     - 3.3.2. Multiple source approach for (eye) artifact correction (MSEC) 19
     - 3.3.3. Surrogate MSEC 20
   - 3.4. Independent component analysis (ICA) 21
     - 3.4.1. Definition of ICA 22
     - 3.4.2. Relation to PCA 23
     - 3.4.3. Statistical independence 24
     - 3.4.4. Contrast functions 25
     - 3.4.5. ICA optimization algorithms 28
     - 3.4.6. Artifact correction by ICA 32
     - 3.4.7. Appendix 33

4. Spatial filter for artifact correction 35
   - 4.1. Basic model 35
   - 4.2. Modeling the artifact subspace 36
   - 4.3. Modeling the signal subspace 37
   - 4.4. Spatial filter 37

5. A new approach: Modeling the signal subspace by preselection 39
   - 5.1. Method 39
     - 5.1.1. Amplitude criterion 39
     - 5.1.2. Correlation criterion 39
     - 5.1.3. Number of eigenvectors 40
   - 5.2. Examples 41
     - 5.2.1. Example 1: Influence of correlation threshold 41
     - 5.2.2. Example 2: Influence of number of eigenvectors 42
   - 5.3. Discussion 45

6. A new approach: Modeling the signal subspace by SCICA 47
   - 6.1. Method 47

6.1.1. SCICA algorithm 47
6.1.2. Parameters of the SCICA algorithm 54
6.2. Examples 57
6.2.1. Applying SCICA to simulated data 58
6.2.2. Applying SCICA to EEG data 62
6.2.3. Applying SCICA to MEG data 67
6.3. Speed of convergence 72
6.4. Discussion 74
6.5. Appendix 75
6.5.1. A1: Expanding an orthonormal basis 75
6.5.2. A2: Perturbation of artifact topographies 76
6.5.3. A3: Linear combination matrix 77
6.5.4. A4: Whitening matrix 79
6.5.5. A5: Orthogonal log-likelihood contrast 80

7. Advantages of SCICA over ICA 81
7.1. SCICA and ICA with automated artifact detection by topography 82
7.1.1. Example 1: Removing blinks and eye movements 82
7.1.2. Example 2: Removing cardiac artifacts and eye movements 83
7.2. SCICA and ICA in the case of temporally dependent activity 86
7.2.1. Example 1: Subepochs of signal activity without artifact 86
7.2.2. Example 2: No subepochs of signal activity without artifact 87

8. Summary and Discussion 91

9. References 95

10. Abbreviations 103

11. Mathematical notation and symbols 105

12. Acknowledgements 107

1. Introduction

Spontaneous and event-related recordings of the electroencephalogram (EEG) and magnetoencephalogram (MEG) can be severely contaminated by artifacts such as eye movements, blinks, cardiac and muscle activity or line noise. During review of spontaneous EEG and MEG recordings artifacts may seriously interfere with the detection and analysis of events of interest. In event-related recordings artifact-contaminated trials often have to be excluded from averaging. This may result in an unacceptable loss of data. Therefore, it has become a well established procedure to try to correct artifacts in the recorded data. The prime goal of artifact correction is to remove artifacts completely without distortion of underlying brain signals of interest. The most promising approaches to artifact correction are spatial filters based on artifact and brain signal topographies such as the spatial filters calculated from the multiple source approach for (eye) artifact correction (MSEC) (Berg and Scherg, 1991a,b, 1994) or independent component analysis (ICA) (Vigário et al., 1998; Jung et al., 1998, 2000a,b). Modeling both artifact and brain signal topographies enables artifact removal without distortion of relevant brain signals. The two approaches, MSEC and ICA, differ in the way artifact and brain signal topographies are estimated.

In MSEC, artifact and brain signal topographies are determined separately. In the first step, artifact topographies are extracted from single or averaged artifacts of the same recording session using principal component analysis (PCA). In the second step, brain signal topographies are determined by spatio-temporal dipole source analysis (Scherg and von Cramon, 1985; Scherg, 1990) in the presence of the predefined artifact topographies. As MSEC requires a dipole source model it is essentially useful for averaged event-related data. Surrogate MSEC (Berg and Scherg, 1991b, 1994) presents an extension to continuous (i.e. spontaneous or event-related) recordings. In surrogate MSEC, a steady unoptimized dipole source configuration is used to model brain activity. If brain and artifact activity are spatially correlated, an unoptimized dipole model may, however, not be sufficient to completely separate brain and artifact activity as will be demonstrated in this thesis.

In ICA, artifact and brain signal topographies are determined together. ICA decomposes the artifact-contaminated data segment of m channels into an equal number of statistically independent waveforms and corresponding topographies. In continuous recordings, brain and artifact activity are in general sufficiently independent to separate into different components. The independent artifact components have to be detected manually. This can be very time-consuming (Jung et al., 2000a).

In this thesis two spatial filter approaches for artifact correction in continuous EEG and MEG recordings are presented that in contrast to MSEC do not depend on the existence of a dipole source model and that unlike ICA require no a posteriori manual identification of artifact components. Comparable to MSEC, artifact topographies are derived in advance from single or averaged artifact prototypes. In order to model brain signal topographies two novel concepts are introduced in this thesis: the preselection approach (Ille et al., 1997) and **s**patially **c**onstrained **ICA** (**SCICA**) (Ille et al., 2001).

In the preselection approach a relevant number of eigenvectors is extracted from an artifact-free subset of the data segment. The subset is obtained by excluding sample vectors from the original data segment that exceed a certain amplitude or correlation with the predefined artifact subspace. The aptness of the estimated brain signal topographies, however, depends crucially on the subjective choice of parameters such as the correlation threshold that is hard to estimate automatically.

The new concept of SCICA offers an alternative way of modeling the brain signal topographies. Unlike preselection, it is based on the whole data segment and is free of parameters that are hard to determine automatically. SCICA uses the prior knowledge about artifact topographies and combines this spatial information with the temporal-statistical strategy of ICA to estimate brain signal topographies. Starting from n a priori determined artifact topographies, the artifact-contaminated data segment of m channels is decomposed iteratively into m-n further components until all waveforms are maximally independent under the spatial constraint. The m-n obtained components represent the brain activity. MSEC and spatial filtering with preselection, on the contrary, use only the spatial information about artifacts while ICA applies solely the temporal assumption of statistically independent waveforms.

In chapter 2 an overview over different types of artifact that may contaminate continuous EEG and MEG recordings is given. In chapter 3 alternative approaches to artifact correction are reviewed. In chapter 4 the spatial filter for artifact correction employed in this thesis is presented. In chapter 5 the preselection approach and its parameters are described. Examples illustrate the good performance of the method but also demonstrate its dependence on a suitable choice of the parameter thresholds. In chapter 6 the SCICA approach and an algorithm to perform the SCICA decomposition are introduced. Using simulated and real EEG/MEG data it is demonstrated that SCICA can remove artifacts without relevant distortion of brain activity. Spatial filtering with preselection and SCICA is applied to spontaneous EEG/MEG segments in this thesis. Applicability to continuous event-related recordings is only considered theoretically. The quality of artifact correction is quantified for simulated data. For real data, artifact correction is evaluated by thorough visual comparison of the EEG/MEG segment before and after spatial filtering. In chapter 7 advantages of SCICA over ICA are demonstrated. The thesis closes with a summary and discussion in chapter 8.

2. Artifacts

Recordings of the EEG and MEG may be heavily contaminated by artifacts. Any electric potential or magnetic field that is not generated by the electrical activity of the brain is considered to be an artifact. Artifacts can be of physiological or non-physiological origin. Physiological artifacts are subject-induced. They are mainly caused by eye movements, blinks, cardiac, or muscle activity. Non-physiological artifacts are due to the recording device, electrodes/sensors, or are externally introduced like, for instance, line noise. In sections 2.1 and 2.2 an overview over physiological and non-physiological artifacts is given.

2.1. Physiological artifacts

A variety of physiological artifacts may be found in continuous EEG and MEG recordings. In the next chapters ocular, cardiac, muscle, respiratory and electro-dermal artifacts are considered. Not all artifacts are of general concern. The most severe contamination is caused by blinks and eye movements. Blink and eye movement artifacts are of high amplitude, interfere with brain activity in the low-frequent delta (1-4 Hz) and theta (4–8 Hz) range and occur in almost any EEG or MEG recording even during sleep or with eyes closed. Cardiac and muscle artifacts are of smaller amplitude, with the exception of cardiac artifacts in MEG, and overlap with brain activity in the high-frequent alpha (8–14 Hz), beta (14–30 Hz) and gamma (30–70 Hz) range. They are not commonly observed but especially cardiac activity may be very distracting. Respiratory or electro-dermal artifacts are in general less relevant for the clinical review of spontaneous EEG/MEG recordings as they are already eliminated when using a standard low-cutoff filter of 1 Hz (Ebner et al., 1999). In event-related studies, on the contrary, where DC recording or long time constants may be necessary to register slow event-related components these types of artifact may be very disturbing.

2.1.1. Ocular artifacts

When the eyes move or blink the potential distribution at the scalp changes. Ocular potentials are maximum in the electro-oculogram (EOG) recorded bipolarly between electrodes located near the eyes. Horizontal eye movements are best recorded in the horizontal EOG (HEOG) with electrodes placed in the left and right outer canthi. Vertical eye movements and blinks can be observed in the vertical EOG (VEOG) between electrodes above and below one eye.

Ocular potentials are also recorded at some distance from the eyes and can thus contaminate EEG and MEG recordings. In Fig. 2.1 a blink and a horizontal eye movement in an EEG segment are depicted. The ocular artifacts are of high amplitude. The blink has its maximum (positive) deflection at electrodes Fp1 and Fp2. Note that negativity is shown upwards. The horizontal eye movement is maximum at right frontal electrodes F8/F10 and minimum at left frontal electrodes F7/F9. While an eye movement may last several hundreds of milliseconds, a blink usually does not exceed 200 to 300 ms. The typical topography of a blink and a horizontal eye movement is shown in the voltage maps of Fig. 2.1.

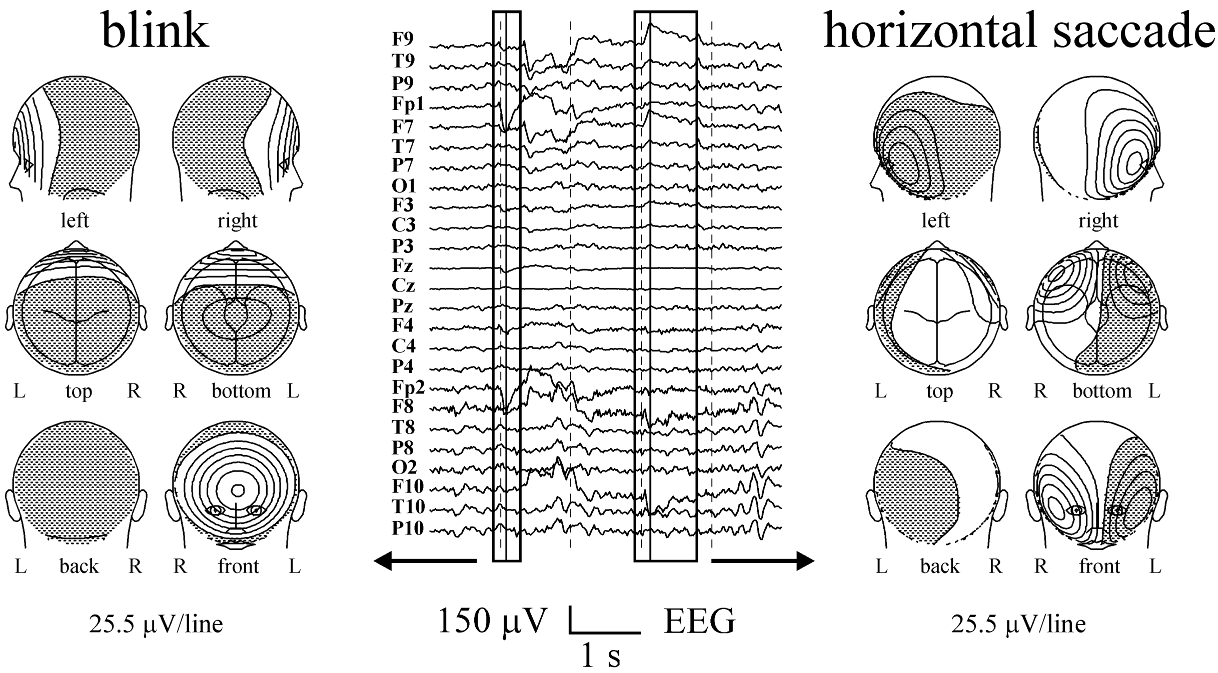

Fig. 2.1: A blink and a horizontal eye movement in an EEG segment (5 s, unfiltered, reference FCz). Negativity is shown upwards. The topography of the ocular artifacts at the marked time point in the boxes is depicted in the voltage maps. Areas of negative potential are hatched in the maps.

The source of the ocular artifacts is the corneo-retinal potential difference. The cornea is charged positively relative to the retina. To describe its electrical and magnetic properties, the eye has often been modeled as a current dipole oriented along the optic axis of the eye (e.g. Barry and Melvill Jones, 1965; Katila et al., 1981; Antervo et al., 1985).

When the eyes move, the orientation of the corneo-retinal dipole changes relative to the electrodes/sensors. The difference dipole $\Delta p = p_2 - p_1$ between the corneo-retinal dipole at the beginning (p_1) and the end (p_2) of the eye movement represents the measured field pattern (e.g. Katila et al., 1981). The electrical potential at the scalp in direction of the eye movement becomes positive, in the opposite direction it is negative. Thus, the horizontal eye movement in Fig. 2.1 was caused by a movement of the eyes to the right.

For the generating mechanism of a blink two different hypothesis exist. On the one hand, it has been assumed that the blink potential is caused by an upward rotation of the eyeball under closed eyelids (e.g. Hector, 1980) comparable to Bell's phenomenon at peripheral facial palsy where the eyeball rotates upwards during an attempted eye closure against a paralyzed lid. On the other hand, Matsuo et al. (1975) have shown that no upward rotation of the eyeball occurs during a blink and that already a simple downward movement of the upper eyelid over the cornea changes the electrical potential at the scalp. The eyelids act as a sliding electrode connecting the forehead to the positively charged anterior pole of the eyeball. The eyelid hypothesis implying different generator mechanisms for vertical eye movements and blinks is also supported by the different field patterns that may be observed for vertical saccades and blinks (e.g. Overton and Shagass, 1969; Picton et al., 2000a,b). In Fig. 2.2 the topography of a blink and an upward saccade is depicted. The blink potential falls off more rapidly from the front to the back of the scalp than the potential caused by the saccade. Dipole modeling yielding dipoles with different locations and orientations for blinks and vertical saccades further confirms the eyelid hypothesis (Berg and Scherg, 1991a,b; Lins et al., 1993).

Vertical saccades may be accompanied by an additional spike-like artifact called the rider artifact. The rider artifact has been shown to be generated by the eyelid during upward movement of the eyeball (Barry and Melvill Jones, 1965; Lins et al., 1993).

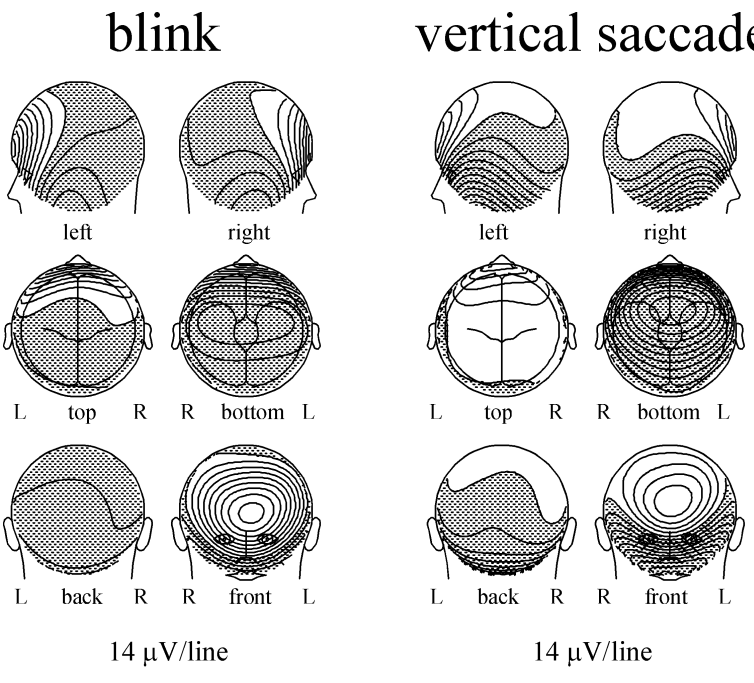

Fig. 2.2: Voltage map of a blink and a vertical upward saccade. The blink potential falls off more rapidly from the front to the back of the scalp.

To avoid ocular artifacts during event-related recordings subjects are often asked to fixate a point or to refrain from eye movement and blinking. However, this introduces a secondary task into the experiment that may interfere with the original task (Brunia et al., 1989).

Moreover, children, psychiatric and neurological patients often cannot adequately follow this instruction. Therefore, EEGs and MEGs may be recorded with eyes closed. This avoids blinking but can introduce eye movements under closed eyelids. Eye artifacts also occur during sleep. They are slow when falling asleep and rapid during the REM (rapid eye movement) sleep stage.

2.1.2. Cardiac artifacts

Recordings of the EEG or MEG (Jousmäki and Hari, 1996) may be contaminated by cardiac artifacts. Usually only the sharp R-wave of the cardiac cycle is observed. Sometimes the T-wave may be visible too. Cardiac artifacts can be identified easily by simultaneously recording the EKG. In the left panel of Fig. 2.3 an EEG segment superimposed by cardiac activity is shown together with the simultaneously recorded EKG. The artifacts are of relatively small amplitude. In the right panel the average of several hundreds of cardiac cycles having an enhanced signal-to-noise ratio (SNR) is depicted. The R-wave is very prominent at temporal electrodes. Even a small T-wave is visible in the average. The topography of the averaged R-wave exhibiting a right temporal negativity and a left temporal positivity is displayed in the voltage map. In MEG recordings cardiac artifacts in temporal sensors may be of considerably higher amplitude than spontaneous brain activity.

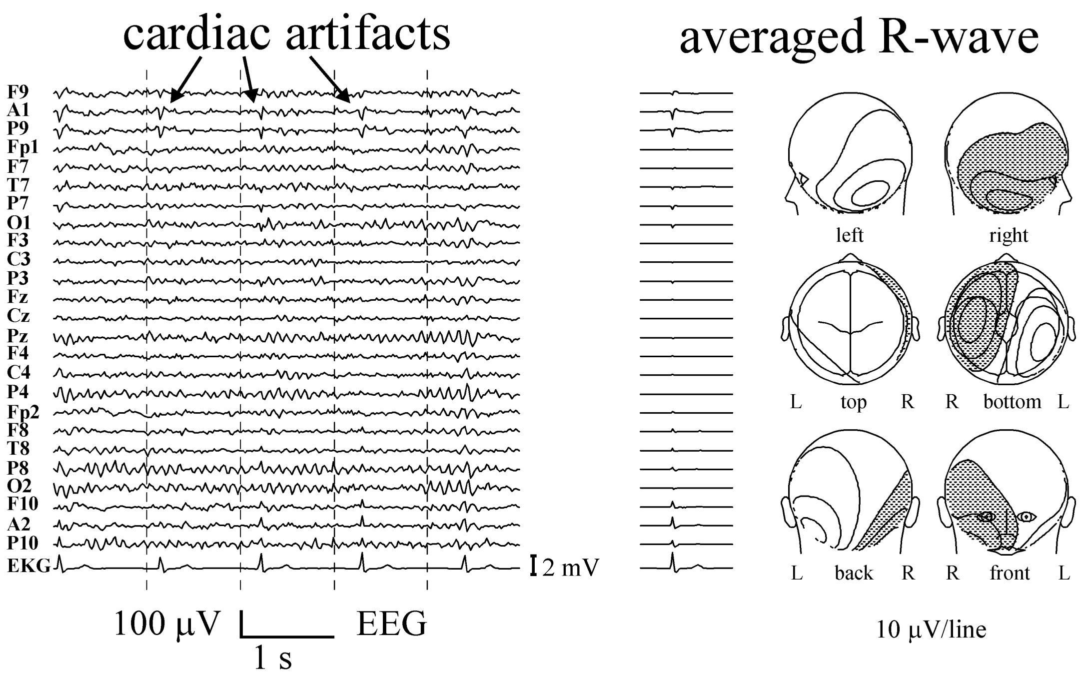

Fig. 2.3: (Left) Cardiac artifacts in an EEG segment (5 s, 1-70 Hz, average-reference). The simultaneously recorded EKG is shown at the bottom. (Right) Averaged cardiac artifact (1 s). The R-wave is most prominent at left and right temporal electrodes. The voltage map depicts the topography of the averaged R-wave.

Cardiac artifacts are caused by the electrical activity of the heart. During a cardiac cycle the potential difference in the heart changes. The location, orientation and magnitude of the heart dipole varies according to Fig. 2.4. At the R-wave, the magnitude of the dipole is maximum. Whether cardiac activity is recorded in the EEG/MEG depends on the individual position of the heart dipole relative to the electrodes/sensors and the particular conductivity properties. Cardiac contamination is more frequent, for instance, in obese or short-necked subjects (Cooper et al., 1984).

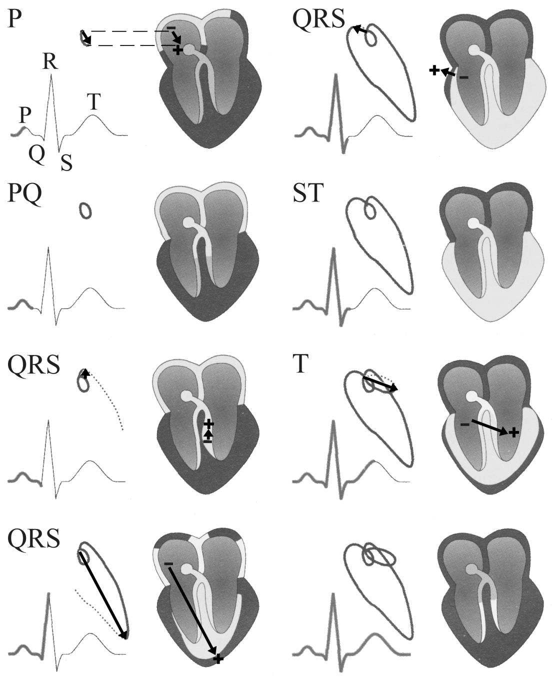

Fig. 2.4: Change of potential differences during a cardiac cycle. The magnitude of the heart dipole reaches its maximum at the R-wave during the QRS-complex. Adapted from Antoni (1993).

In MEG even ballistocardiographic artifacts have been reported (Hari, 1993). They are caused by the movement of magnetic particles, e.g. in clothes, at the rhythm of the heart beat.

2.1.3. Muscle artifacts

When a subject is anxious or not relaxed, high-frequent muscle activity may appear in EEG and MEG recordings (cf. Fig. 2.5). The artifact is caused by the electrical activity of a contracted muscle. Each muscle is composed of a great number of 'motor units' which are activated proportionately to the strength of the muscular contraction. On the one hand, isolated muscle potentials may occur, for instance rectus lateralis spikes during photo stimulation. On the other hand, complex high-frequent interference patterns (cf. Fig. 2.5) may be observed if many motor units fire at high frequencies.

Muscle artifacts are mainly recorded in electrodes or sensors over the contracted muscle. They may be seen at any recording position, but predominantly occur over temporal regions, for instance, if subjects contract their jaws or clench their teeth (Hector, 1980). As a remedy subjects may be asked to open the mouth (Cooper et al., 1984). Frontal muscle artifacts, for example, may be caused by frowning. Occipital muscle activity occurs if the subject's head is badly positioned.

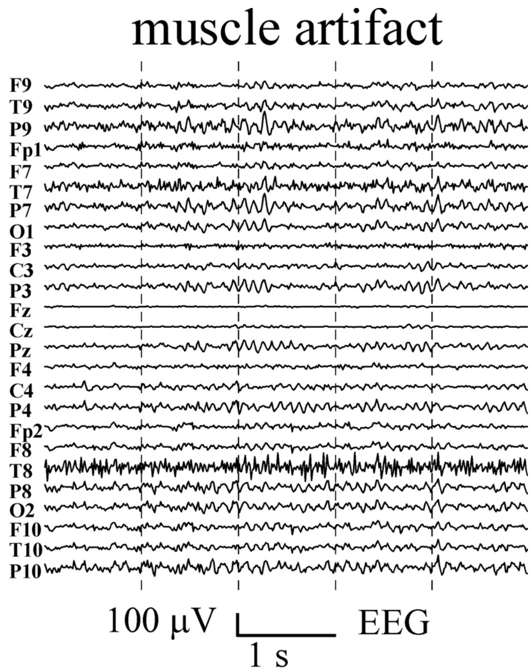

Fig. 2.5: An EEG segment (5 s, 1-70 Hz, referenced to FCz) contaminated by high-frequent muscle activity at right temporal electrode T8.

2.1.4. Other physiological artifacts

Sometimes EEG or MEG recordings are contaminated by respiratory artifacts. In EEG, the contamination is caused by the mechanical displacement of an electrode during respiration. Respiratory artifacts are very frequent in children, especially during hyperventilation (Hector, 1980). The artifact occurs synchronously with respiration, i.e. every 2-6 s for adults and faster in infants. In MEG recordings, this type of artifact is only observed if magnetic material, e.g. in clothes, is moved by respiration.

When the skin resistance changes electro-dermal activity may be recorded in the EEG. Electro-dermal artifacts are very slow waves of high amplitude (Hector, 1980). The skin resistance is mainly influenced by sweating, but also by instructions or noise.

2.2. Non-physiological artifacts

Besides subject-induced physiological artifacts a couple of non-physiological artifacts caused, for instance, by the recording device, electrodes/sensors, or cables may be found in continuous EEG and MEG recordings. In EEG recordings, the 50 Hz mains frequency may be introduced as line noise into one or several electrodes (cf. Fig. 2.6). The artifact can be eliminated with a 50 Hz notch filter unless high-frequency phenomena in the gamma band are of interest. In some recordings, there may also be high-amplitude electrode artifacts caused, for example, by the movement of a badly fixed electrode (cf. Fig. 2.7).

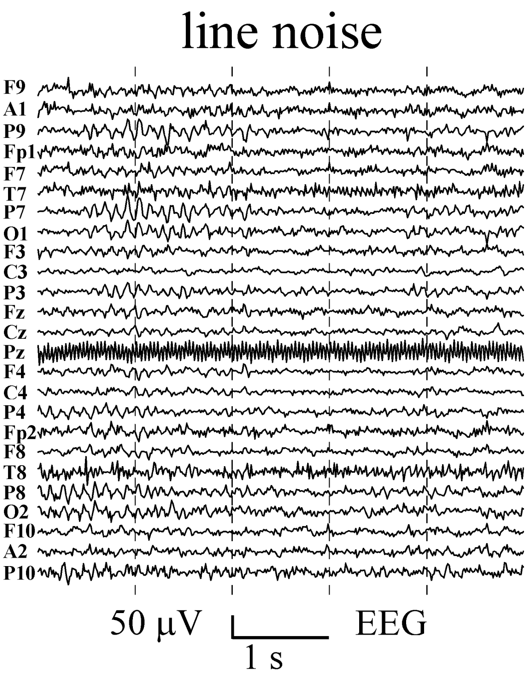

Fig. 2.6: 50 Hz line noise at electrode Pz in an EEG segment (5 s, 1-70 Hz, average-reference).

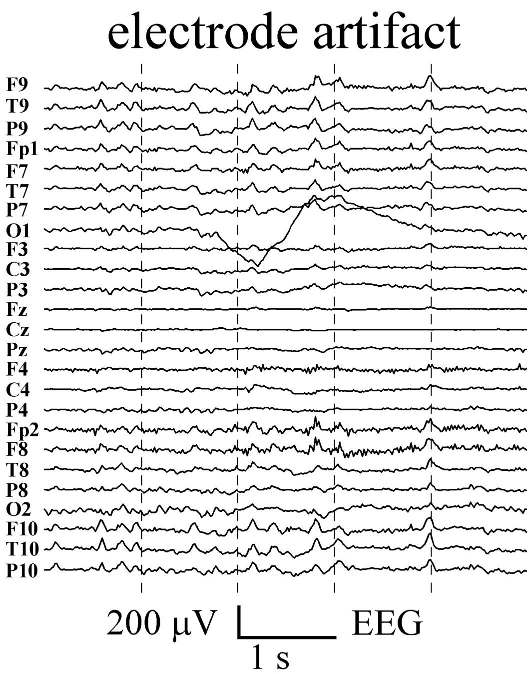

Fig. 2.7: An electrode artifact at electrode O1 in an EEG segment (5 s, 0.1-70 Hz, reference FCz).

In MEG recordings, environmental artifacts are reduced to some extent by carrying out the measurement in a magnetically shielded room. Gradiometer coils or compensation coils in magnetometer systems further diminish artifact contamination. Nevertheless, external magnetic fields may be a severe problem in MEG recordings. Non-physiological contamination may also occur inside the shielded room, e.g. by a digital watch (Hari, 1993). Under suitable recording conditions, however, non-physiological artifacts in EEG and MEG recordings can be largely avoided.

3. Existing methods for artifact correction

In the previous chapter different types of artifact that may contaminate spontaneous or event-related EEG and MEG recordings have been discussed. Using a standard filter setting of 1-70 Hz for clinical review of spontaneous data (Ebner et al., 1999) some artifacts are already eliminated. Artifacts such as eye movement, blink, or cardiac activity, however, that interfere with the frequency range of relevant brain activity cannot be removed simply by temporal filtering and may be found very distracting during review and analysis of a spontaneous EEG/MEG. In event-related recordings, artifact-contaminated trials are often excluded from averaging. This may result in an unacceptable loss of data. Therefore, it has become a well established procedure to try to correct artifacts in the recorded data. The main goal of artifact correction is to remove artifacts completely but largely preserve the topography of the underlying brain signals of interest. In this chapter the major approaches to artifact correction are presented and evaluated: EOG subtraction, projection method, multiple source eye correction (MSEC), and independent component analysis (ICA). All methods except EOG subtraction can be expressed as spatial filters. While the spatial filter in the projection method is based on artifact topographies only, the spatial filters derived from MSEC and ICA consider both artifact and brain signal topographies and, thus, enable a distortion-free separation of artifact and brain activity.

3.1. EOG subtraction

EOG subtraction methods (reviewed e.g. by Jervis et al., 1988 and Brunia et al., 1989) deal primarily with the removal of ocular artifacts from event-related EEG recordings. For each EEG channel the proportion of ocular contamination is estimated utilizing one or several simultaneously recorded EOG derivations as reference channels. In order to yield the corrected data the EOG signals are scaled by the estimated proportion and subtracted from the original EEG signals. In the so-called time domain approach the scaling factor is equal for each frequency as opposed to the frequency domain approach where the scaling function varies with frequency. In chapter 3.1.1 the correction formulae of the time and frequency domain approach are contrasted. In chapter 3.1.2 different techniques to estimating the scaling factors and functions are collected. In 3.1.3 subtracting the EOG itself is addressed. It is shown that subtracting the EOG may distort brain activity as the EOG is no clean artifact channel but may contain brain activity as well.

3.1.1. Time and frequency domain approach

A wealth of EOG subtraction variants has been proposed using either on-line analogous circuits (e.g. Girton and Kamiya, 1973; Barlow and Rémond, 1981) or off-line correction formulae in the time or frequency domain (time domain, e.g. Hillyard and Galambos, 1970; Corby and Kopell, 1972; Verleger et al., 1982; Gratton et al., 1983; Fortgens and de Bruin, 1983; Elbert et al., 1985; Schwind and Dormann, 1986; Semlitsch et al., 1986; Van den Berg-Lenssen et al., 1989; Croft and Barry, 2000; frequency domain, e.g. Whitton et al., 1978; Woestenburg et al., 1983; Gasser et al., 1985, 1986). The time domain approach to EOG subtraction can be summarized in the following correction formula:

$$\mathrm{d}_{\mathrm{corr},i}(t) = \mathrm{d}_i(t) - \sum_j r_{ij}\, \mathrm{d}_{\mathrm{EOG},j}(t)$$

where $\mathrm{d}_{\mathrm{EOG},j}(t)$ and $\mathrm{d}_i(t)$ are the signals measured in the j^{th} EOG channel and in the i^{th} EEG channel. The scaling factor r_{ij}, also referred to as propagation factor, describes the relationship between the i^{th} EEG channel and the j^{th} EOG channel, $\mathrm{d}_{\mathrm{corr},i}(t)$ is the corrected EEG signal. The frequency domain approach takes into account a frequency-specific relationship between EOG and EEG. It may be described by the following correction formula:

$$\mathrm{d}_{\mathrm{corr},i}(t) = \mathrm{d}_i(t) - \sum_j \mathcal{F}^{-1}\left\{\mathrm{r}_{ij}(f)\, \underline{\mathrm{d}}_{\mathrm{EOG},j}(f)\right\}$$

where $\underline{\mathrm{d}}_{\mathrm{EOG},j}(f)$ is the complex Fourier transform of the j^{th} EOG signal. The generally complex scaling function $\mathrm{r}_{ij}(f)$ (also called transfer function) describes the relationship between the i^{th} EEG channel and the j^{th} EOG channel for each frequency *f*. The operator $\mathcal{F}^{-1}$ denotes the inverse Fourier transformation. Thus, in the frequency domain approach each EOG signal is filtered before subtraction using its individual scaling function.

3.1.2. Scaling factors and functions

In order to estimate the scaling factors/functions different techniques are applied, e.g. calculation of amplitude ratios (e.g. Hillyard and Galambos, 1970; Corby and Kopell, 1972) respectively ratios between frequency spectra (Whitton et al., 1978) or linear regression (time domain, e.g. Verleger et al., 1982; Gratton et al., 1983; frequency domain, e.g. Woestenburg et al., 1983; Gasser et al., 1985, 1986). Different numbers and types of EOG derivations are propagated to comprise all directions of eye movements (e.g. Fortgens and de Bruin, 1983; Elbert et al., 1985; Jervis et al., 1988). The scaling factors are either determined in calibration data sets containing voluntary eye movements and a negligible portion of brain activity only (e.g. Hillyard and Galambos, 1970; Fortgens and de Bruin, 1983) or are estimated directly in the experiment to bypass possible differences between voluntary and involuntary eye movements (e.g. Verleger et al., 1982; Gratton et al., 1983). Calculating the scaling factor in the experiment entails the necessity to consider brain activity in EEG or EOG (e.g. Gratton et al., 1983; Elbert et al., 1985; Gasser et al., 1985, 1986; Schwind and Dormann, 1986) in order to minimize errors in the scaling factor estimates. Brain activity is taken into account, e.g. by incorporating the average of selected artifact-free trials as a further regressor in the multiple regression model (Schwind and Dormann, 1986) or by correcting the scaling function for coherent brain activity in the EOG (Gasser et al., 1985). Coherent brain activity is estimated from selected artifact-free trials too. These approaches are only approximate as they do not attempt to model brain activity that actually underlies an individual trial. Moreover, they depend on recognizing artifact-free epochs of brain activity in the recorded data. Low-frequent brain activity in EOG channels, for instance, that is hardly distinguishable from ocular artifacts, will therefore most probably not be taken into account.

3.1.3. Subtracting the EOG

However, the major problem of EOG subtraction arises during subtraction itself. On the one hand, subtraction significantly reduces the ocular artifacts up to small residuals (e.g. Berg, 1986; Iacono and O'Toole, 1987; Lins et al., 1993). On the other hand, subtraction can distort the topography of relevant brain signals if the subtracted EOG is not free of brain activity (e.g. Berg, 1986; Brunia et al., 1989; Berg and Scherg, 1994; Jung et al., 2000b). To partially avoid distortion, the EOG signals may be filtered with a low pass of 8 Hz before subtraction (Whitton et al., 1978; Jervis et al., 1988). Yet, this will not remove slow brain waves from the EOG. Theoretically, filtering the EOG with an individual scaling function corrected for coherent brain activity (Gasser et al., 1985) is an adequate approach. Practically, the problem of estimating coherent brain activity remains unsolved by EOG subtraction techniques.

3.2. Projection method

The projection method was originally developed by Ilmoniemi (1992) and was applied to artifact correction, e.g. by Huotilainen et al. (1995). In chapter 3.2.1 the method is described and evaluated. It is shown that artifact correction that is based on artifact topographies only results in a distortion of brain signal topographies. In chapter 3.2.2 the projection method is summarized in a spatial filter operator.

3.2.1. Projection method for artifact correction

In the projection method for each topography $\mathbf{d}_i \in \Re^m$ of the original data segment $\mathbf{D} = (\mathbf{d}_1, \ldots, \mathbf{d}_t) \in \Re^{m \times t}$ the portion in direction of a predefined and normalized artifact topography $\mathbf{a} \in \Re^m$ is projected out according to

$$\mathbf{D}_{\text{corr}} = \mathbf{D} - \mathbf{a}\left(\mathbf{a}^{\text{T}}\mathbf{D}\right)$$

leaving the corrected data matrix $\mathbf{D}_{\text{corr}} = \left(\mathbf{d}_{\text{corr},1}, \ldots, \mathbf{d}_{\text{corr},t}\right) \in \Re^{m \times t}$ of m channels and t time points. The unit vector **a** models the artifact. In Huotilainen et al. (1995) **a** is obtained from a measured artifact epoch as the normalized topography $\mathbf{d}_{\mathbf{a},t_{\max}} \in \Re^m$ at the time point $t_{\max}$ of maximum amplitude:

$$\mathbf{a} = \frac{\mathbf{d}_{\mathbf{a},t_{\max}}}{\left\|\mathbf{d}_{\mathbf{a},t_{\max}}\right\|}.$$

Note that $\mathbf{a}^{\text{T}}\mathbf{D}$ in the first equation yields the time course of the predefined artifact topography and that $\mathbf{a}\left(\mathbf{a}^{\text{T}}\mathbf{D}\right)$ is the estimated artifact activity $\hat{\mathbf{D}}_{\text{a}} = \left(\hat{\mathbf{d}}_{\text{a},1}, \ldots, \hat{\mathbf{d}}_{\text{a},t}\right) \in \Re^{m \times t}$ at each electrode/sensor.

In panel (A) of Fig. 3.1 the projection method is applied to a topography $\mathbf{d}_i \in \Re^2$. The estimated artifact topography $\hat{\mathbf{d}}_{\text{a},i}$, i.e. the portion of $\mathbf{d}_i$ in direction of **a**, is obtained by orthogonal projection of $\mathbf{d}_i$ onto **a**. Subtracting $\hat{\mathbf{d}}_{\text{a},i}$ from $\mathbf{d}_i$ yields the corrected topography $\mathbf{d}_{\text{corr},i}$ that is perpendicular to **a**. In panels (B)-(D) it is illustrated why brain signal topographies are generally distorted by the projection method. In panel (B) it is first assumed that the measured vector $\mathbf{d}_i$ is the sum of artifact topography $\mathbf{d}_{\text{a}}$ and an orthogonal brain signal vector $\mathbf{d}_{\text{s}}$. Only in this particular case $\hat{\mathbf{d}}_{\text{a},i}$ and $\mathbf{d}_{\text{corr},i}$ are exact estimates of $\mathbf{d}_{\text{a}}$ and $\mathbf{d}_{\text{s}}$. If $\mathbf{d}_{\text{a}}$ and $\mathbf{d}_{\text{s}}$ are not orthogonal as in panel (C), the artifact is still completely eliminated but as the brain signal topography is not taken into account the portion of $\mathbf{d}_{\text{s}}$ in direction of **a** is mistaken for artifact activity and $\hat{\mathbf{d}}_{\text{a},i}$ is overestimated. Consequently, $\mathbf{d}_{\text{corr},i}$ is a distorted estimate of $\mathbf{d}_{\text{s}}$. The distortion is even worse in panel (D) where $\mathbf{d}_i$ is equal to $\mathbf{d}_{\text{s}}$ and no artifact at all is present. The distortion increases as the angle between $\mathbf{d}_{\text{s}}$ and **a** decreases. The depicted angle of distortion is equal to $90° - \sphericalangle(\mathbf{a}, \mathbf{d}_{\text{s}})$. Distortion can only be avoided if brain signal topographies are taken into account by artifact correction.

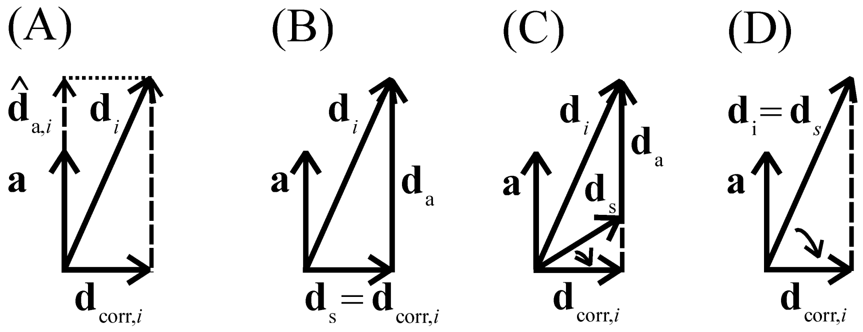

Fig. 3.1: Applying the projection method to the measured vector $\mathbf{d}_i$. (A) The vector $\hat{\mathbf{d}}_{\mathrm{a},i}$, i.e. the portion of $\mathbf{d}_i$ in direction of the predefined normalized artifact topography $\mathbf{a}$, is subtracted from $\mathbf{d}_i$ leaving the corrected vector $\mathbf{d}_{\mathrm{corr},i}$. (B)-(D) Only if $\mathbf{d}_i$ is the sum of artifact topography $\mathbf{d}_\mathrm{a}$ and an orthogonal signal topography $\mathbf{d}_\mathrm{s}$ as shown in panel (B), $\hat{\mathbf{d}}_{\mathrm{a},i}$ and $\mathbf{d}_{\mathrm{corr},i}$ are correct estimates of $\mathbf{d}_\mathrm{a}$ and $\mathbf{d}_s$. Otherwise the subtracted artifact portion is overestimated resulting in a distortion of $\mathbf{d}_s$. The angle of distortion is depicted in panels (C) and (D).

3.2.2. Spatial filter

The projection method can be summarized in the spatial filter operator $\mathbf{F} \in \Re^{m \times m}$

$$\mathbf{F} = \mathbf{I} - \mathbf{a}\mathbf{a}^\mathrm{T}$$

with identity matrix $\mathbf{I} \in \Re^{m \times m}$ or in

$$\mathbf{F} = \mathbf{I} - \mathbf{A}\mathbf{A}^\dagger$$

in the more general case of n artifact topographies $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_n) \in \Re^{m \times n}$. As is apparent from the first equation in the previous subsection the corrected data $\mathbf{D}_\mathrm{corr}$ are obtained by premultiplying the spatial filter $\mathbf{F}$ to the data matrix $\mathbf{D}$. The spatial filter completely removes the modeled artifacts, but equally suppresses any spatially correlated brain activity as it is composed of artifact topographies only and does not take into account brain activity.

3.3. Multiple source eye correction (MSEC)

Multiple source eye correction (MSEC) (Berg and Scherg, 1991a,b, 1994) was introduced for the removal of ocular artifacts from averaged event-related data, but can be easily extended to the correction of other types of artifact. MSEC is based on the spatio-temporal multiple source approach (Scherg and von Cramon, 1985; Scherg, 1990). Therefore, in chapter 3.3.1 the multiple source approach is briefly introduced. In 3.3.2 artifact correction by MSEC is described. It is assessed that artifact correction taking into account artifact and brain activity preserves brain activity that is otherwise distorted. In chapter 3.3.3 a variant of MSEC called surrogate MSEC is presented. Surrogate MSEC refrains from exact modeling of brain activity and is, thus, in principle also applicable to continuous data. Using a segment of continuous data it is illustrated, however, that approximate modeling of brain activity may not be sufficient if artifact and brain activity are spatially correlated.

Artifact correction by MSEC can also be expressed as a spatial filter. In contrast to the projection method the spatial filter is based on artifact and brain signal topographies. The spatial filter operator is derived in chapter 4. It is also used in the novel approaches to artifact correction introduced in this thesis.

3.3.1. Multiple source approach

In the multiple source approach (Scherg and von Cramon, 1985; Scherg, 1990; Mosher et al., 1992) the activity $\mathbf{D} \in \Re^{m \times t}$ measured in the EEG or MEG at m electrodes/sensors and t time points is assumed to be generated by superposition of l^* unknown dipole sources:

$$\mathbf{D} = \mathbf{C}^* \mathbf{S}^* .$$

The column vector $\mathbf{c}_i^* \in \Re^m$ of $\mathbf{C}^* = \left(\mathbf{c}_1^*, \ldots, \mathbf{c}_{l^*}^*\right) \in \Re^{m \times l^*}$ represents the location and orientation of the i^{th} dipole source. The matrix $\mathbf{S}^* = \left(\mathbf{s}_1^*, \ldots, \mathbf{s}_{l^*}^*\right)^{\mathrm{T}} \in \Re^{l^* \times t}$ contains the source waveforms with $\mathbf{s}_i^* \in \Re^t$ corresponding to dipole topography $\mathbf{c}_i^*$. Dipole source modeling of data matrix $\mathbf{D} \in \Re^{m \times t}$, e.g. a segment of averaged evoked data, can be summarized in:

$$\left\|\mathbf{D} - \mathbf{C}\mathbf{S}\right\|_2 = \min$$

where $\mathbf{C} = \left(\mathbf{c}_1, \ldots, \mathbf{c}_l\right) \in \Re^{m \times l}$ is a matrix of l estimated dipole topographies ($l \leq m$). Each topography $\mathbf{c}_i \in \Re^m$ is generated by a dipole with a particular orientation and location in the underlying head model. The matrix $\mathbf{S} \in \Re^{l \times t}$ containing the l corresponding source waveforms is determined according to

$$\mathbf{S} = \mathbf{C}^{\dagger} \mathbf{D}$$

where $\mathbf{C}^{\dagger} \in \Re^{l x m}$ is the Moore-Penrose pseudo-inverse of **C**. Starting from an initial solution dipole orientations and locations are adjusted iteratively using the simplex algorithm for non-linear optimization (Nelder and Mead, 1965) until the best least-square fit to the data is achieved. The optimization may be subject to further constraints such as left-right symmetry of dipoles. Alternatively, source locations may be derived from the individual anatomy or other imaging techniques, e.g. functional magnetic resonance imaging (fMRI) (Scherg et al., 1999).

3.3.2. Multiple source approach for (eye) artifact correction (MSEC)

The multiple source approach for (eye) artifact correction (MSEC) is a special case of the multiple source approach. In MSEC, the data segment $\mathbf{D} \in \Re^{m \times t}$ is regarded as the sum of artifact activity $\mathbf{D}_{\mathrm{A}} \in \Re^{m \times t}$ and brain activity $\mathbf{D}_{\mathrm{B}} \in \Re^{m \times t}$:

$$\mathbf{D} = \mathbf{D}_{\mathrm{A}} + \mathbf{D}_{\mathrm{B}} = \mathbf{A}^{*}\,\mathbf{S}_{\mathrm{A}*} + \mathbf{B}^{*}\,\mathbf{S}_{\mathrm{B}*}$$

where $\mathbf{D}_{\mathrm{A}}$ is assumed to be generated by the superposition of n^* unknown artifact sources and $\mathbf{D}_{\mathrm{B}}$ is postulated to be the linear sum of p^* unknown brain dipole sources. The columns of $\mathbf{A}^{*} \in \Re^{m \times n^*}$ and $\mathbf{B}^{*} \in \Re^{m \times p^*}$ represent the topographies of the artifact and brain sources. The rows of $\mathbf{S}_{\mathrm{A}*} \in \Re^{n^* \times t}$ and $\mathbf{S}_{\mathrm{B}*} \in \Re^{p^* \times t}$ contain the corresponding waveforms.

First, a set of not necessarily dipolar topographies $\mathbf{A} \in \Re^{m \times n}$ is determined yielding an estimate of the artifact subspace $\mathrm{range}(\mathbf{A}^{*})$. Ocular artifact topographies, for instance, may be extracted from calibration data with eye movements and blinks using PCA. Ocular artifacts are not modeled by dipoles located somewhere in both eyes as the head model is usually not accurate enough around the eyes. Estimating the artifact subspace is further addressed in chapter 4.2.

Dipole source modeling of brain activity is performed in the presence of the predefined artifact topographies $\mathbf{A} \in \Re^{m \times n}$:

$$\left\| \mathbf{D} - \mathbf{A}\mathbf{S}_{\mathrm{A}} - \mathbf{B}\mathbf{S}_{\mathrm{B}} \right\|_2 = \min$$

where $\mathbf{B} \in \Re^{m \times p}$ is the matrix of p dipolar brain signal topographies ($n + p \le m$). The matrices $\mathbf{S}_{\mathrm{A}}$ and $\mathbf{S}_{\mathrm{B}}$ contain the corresponding waveforms:

$$\begin{pmatrix} \mathbf{S}_{\mathrm{A}} \\ \mathbf{S}_{\mathrm{B}} \end{pmatrix} = (\mathbf{A}\ \mathbf{B})^{\dagger}\, \mathbf{D}$$

where $(\mathbf{A}\ \mathbf{B})^{\dagger}$ denotes the Moore-Penrose pseudo-inverse of the compound matrix $(\mathbf{A}\ \mathbf{B}) \in \Re^{m \times (n+p)}$. The predefined artifact topographies remain unchanged during optimization of the brain signal dipoles. Only their waveforms vary depending on the current brain signal topographies. Complete separation of artifact and brain signal activity is achieved if the estimated artifact signals are free of brain activity and vice versa. The corrected data matrix $\mathbf{D}_{\mathrm{corr}} \in \Re^{m \times t}$ may finally be obtained by subtracting the estimated artifact activity at the electrodes/sensors ($\mathbf{A}\mathbf{S}_{\mathrm{A}}$) from the original data matrix:

$$\mathbf{D}_{\mathrm{corr}} = \mathbf{D} - \mathbf{A}\mathbf{S}_{\mathrm{A}}\,.$$

The MSEC approach is applicable to averaged event-related recordings or signals of sufficient SNR, for example epileptiform spikes (Ebersole, 1991), provided that a dipole source model exists for the data under investigation. Comparison of MSEC with traditional correction by EOG subtraction has confirmed that incorporating brain signal dipoles preserves brain activity especially in frontal channels that is otherwise removed (Berg and Scherg, 1991b, 1994).

3.3.3. Surrogate MSEC

For continuous data in the absence of suitable optimized dipole source models, Berg and Scherg (1991b, 1994) have proposed the surrogate MSEC. In surrogate MSEC, a set of dipole sources is placed at strategic positions of the brain. These dipoles are not optimized and are expected to explain some or most of the brain activity. By definition, surrogate MSEC is less effective than the optimizing variant of MSEC as the estimate of brain activity is only approximate. However, for averaged event-related data surrogate MSEC has been shown to still perform better than traditional correction by EOG subtraction (Berg and Scherg, 1994). In Fig. 3.2 surrogate MSEC is applied to continuous EEG data. The example illustrates that an unoptimized dipole model may not model spatially correlated brain and artifact activity adequately. The original EEG segment (5 s, 1-70 Hz, referenced to the average of F3 and F4) shown in the left panel contains a blink in the 1^{st} second and an oblique eye movement in the 2^{nd} second followed by rhythmic brain activity. Brain and artifact activity occur in similar channels (e.g. C3, P3, P4, O1, O2, P7, Pz, P9) suggesting their correlation. In the right panel the surrogate MSEC model is displayed. Sources 1-4 are the unoptimized compensation sources used to explain brain activity. They are so-called regional sources consisting of 3 orthogonal dipoles at the same location. Therefore, for each regional source an overplot of three waveforms is displayed. The position of the compensation sources was adopted from Berg and Scherg (1991b, 1994). Sources 5 and 6 model artifact activity. Their topographies are derived from a PCA over the first two seconds of the original data segment containing mainly artifact activity. Component 5 explains 84 %, component 6 still 9 % of the variance. For display in the head schemes the PCA topographies are represented by their best-fit dipoles yielding a center of gravity. Using sources 1-5 a considerable amount of artifact still projects onto brain sources 1 and 4. If this model was applied, artifact correction would be incomplete. Adding artifact component 6 all brain signal waveforms are free of artifact activity. Component 6, however, represents a considerable amount of brain activity. Applying the latter model would distort correlated brain activity.

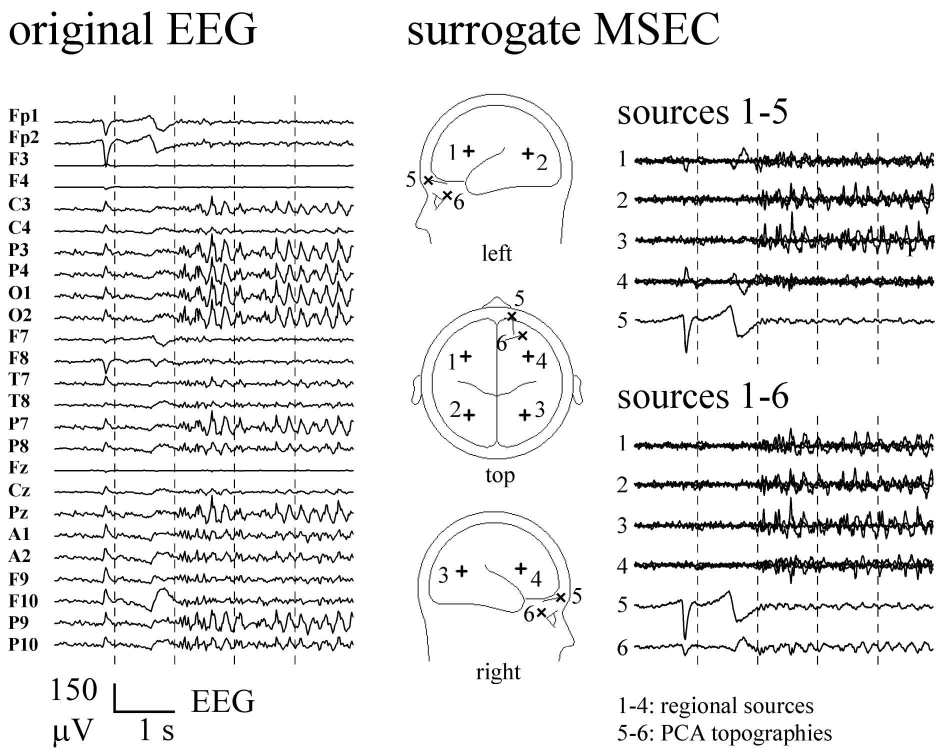

Fig. 3.2: Applying surrogate MSEC to a continuous EEG segment. (Left) Original EEG segment (5 s, 1-70 Hz, reference: average of F3/F4) with spatially correlated artifact and brain signal activity. (Right) The surrogate MSEC model. Regional sources 1-4 are placed at strategic positions in the brain. They are supposed to compensate for brain activity. PCA topographies 5 and 6 model artifact activity. PCA has been calculated over the first two seconds of the original segment. Either model (sources 1-5/1-6) results in an incomplete separation of artifact and brain signal activity.

3.4. Independent component analysis (ICA)

Most recently ICA has been suggested and successfully employed for artifact correction (Makeig et al., 1996; Vigário, 1997; Vigário et al., 1998; Jung et al., 1998, 2000a,b). ICA decomposes a data segment of *m* channels into an equal number of topographies and statistically independent waveforms. The independent components are achieved by optimizing suitable contrast (objective) functions. In chapter 3.4.1 the problem of ICA is defined. In 3.4.2 the relation of ICA to PCA is discussed. The term statistical independence is illustrated in section 3.4.3 and distinguished from the condition of uncorrelatedness that plays a key role in PCA. In 3.4.4 some important contrast functions are presented. Their close relationship is pointed out. Optimization of the contrast functions is treated in chapter 3.4.5 focusing on the extended infomax (Lee et al., 1999) and JADE algorithms (Cardoso and Souloumiac, 1993) that will be employed for comparison with the new SCICA algorithm in chapter 7. Extended infomax is also referred to as extended ICA in the literature. JADE stands for joint approximate diagonalization of eigenmatrices. In chapter 3.4.6 application of ICA to artifact correction is summarized. It is assessed that manual identification of independent components representing artifact activity can be time-consuming especially if an artifact is spread over different independent components. Artifact correction by ICA can be expressed by the same spatial filter operator as MSEC. The main difference between both approaches is how artifact and brain signal topographies are estimated. The spatial filter operator itself that will be further employed in this thesis is derived in chapter 4. Finally, in the Appendix in chapter 3.4.7 the terms kurtosis and cumulants referred to in the text are defined. Further overviews of ICA may be found, for instance, in Cardoso (1998), Hyvärinen (1999), or Hyvärinen and Oja (2000).

3.4.1. Definition of ICA

The basic assumption of linear ICA is that the observed data segment $\mathbf{D} = (\mathbf{d}_1, \ldots, \mathbf{d}_m)^{\mathrm{T}} \in \Re^{m \times t}$ of m channels and t time points is the weighted sum of m statistically independent waveforms $\mathbf{S}^* = \left(\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*\right)^{\mathrm{T}} \in \Re^{m \times t}$:

$$\mathbf{D} = \mathbf{C}^* \, \mathbf{S}^*$$

where $\mathbf{C}^* \in \Re^{m \times m}$ denotes the unknown mixing matrix that is supposed to have full column rank. The i^{th} column vector of $\mathbf{C}^*$ describes the topography of the i^{th} waveform $\mathbf{s}_i^*$. Each of the t samples of the column vector $\mathbf{s}_i^*$ is assumed to be an independent realization of the random variable s_i^* distributed with probability density function (pdf) $\mathrm{p}_i(s_i^*)$. The vectors $\mathbf{s}_i^*$ are expected to have zero mean, i.e. $\mathrm{E}\{\mathbf{s}_i^*\}=0$. Centered vectors $\mathbf{s}_i^*$ can be easily achieved by subtracting the mean $\overline{m}_i$ from each observed waveform, i.e. $\mathrm{E}\{\mathbf{d}_{\mathrm{i}}\}=0$. In the rest of chapter 3.4 when talking about the (observed) data segment $\mathbf{D}$ it is assumed that $\mathrm{E}\{\mathbf{d}_{\mathrm{i}}\}=0$.

The problem of ICA is to find a linear transformation or unmixing matrix $\mathbf{W} \in \Re^{m \times m}$ so that the corresponding waveforms

$$\mathbf{S} = \mathbf{W} \, \mathbf{D}$$

are an estimate of $\mathbf{S}^*$. In order to solve the problem ICA uses information on the distribution of the unknown independent waveforms. The information is either a priori available or is estimated from the data. Comon (1994) has shown that it is possible to extract the independent waveforms from the received mixtures if at most one of the independent waveforms is normally distributed.

Note that the initially subtracted mean can easily be considered after estimating the unmixing matrix by adding $\mathbf{WM}$ back to the centered estimate $\mathbf{S}$ with $\mathbf{M} = (\mathbf{m}_1, \ldots, \mathbf{m}_t) \in \Re^{m \times t}$ and $\mathbf{m}_i = (\overline{m}_1, \ldots, \overline{m}_m)^{\mathrm{T}}$ for i=1...t.

3.4.2. Relation to PCA

In this chapter the relation between ICA and PCA is established. Both, ICA and PCA, are linear transformation techniques. PCA decomposes the observed data segment $\mathbf{D} \in \Re^{m \times t}$ into $r = \mathrm{rank}(\mathbf{D})$ topographies $\overline{\mathbf{C}} \in \Re^{m \times r}$ and waveforms $\overline{\mathbf{S}} = (\overline{\mathbf{s}}_1, \ldots, \overline{\mathbf{s}}_r)^{\mathrm{T}} \in \Re^{r \times t}$:

$$\mathbf{D} = \overline{\mathbf{C}}\,\overline{\mathbf{S}}\,.$$

The columns of the topography matrix $\overline{\mathbf{C}}$ are equal to the eigenvectors of $\mathbf{D}\mathbf{D}^{\mathrm{T}}$. Thus, they are orthonormal, i.e.

$$\overline{\mathbf{C}}^{\mathrm{T}}\,\overline{\mathbf{C}} = \mathbf{I}$$

in contrast to ICA where the independent topographies $\mathbf{C}$ are usually not orthogonal. The basic waveforms $\overline{\mathbf{s}}_i$ are decorrelated in time:

$$\mathrm{E}\left\{\overline{\mathbf{s}}_i^{\mathrm{T}}\overline{\mathbf{s}}_j\right\} - \mathrm{E}\left\{\overline{\mathbf{s}}_i\right\}\mathrm{E}\left\{\overline{\mathbf{s}}_j\right\} = 0 \quad \text{for } i \neq j\,.$$

Note that the basic waveforms have zero mean as the observed waveforms have been initially centered. PCA may be calculated as the singular value decomposition (SVD) of $\mathbf{D}$

$$\mathbf{D} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^{\mathrm{T}}$$

with left singular vectors $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_r) \in \Re^{m \times r}$, singular values $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r) \in \Re^{r \times r}$, $\sigma_1 \geq \ldots \geq \sigma_r$, right singular vectors $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_r) \in \Re^{t \times r}$, $\overline{\mathbf{C}} = \mathbf{U}$ and $\overline{\mathbf{S}} = \mathbf{\Sigma}\,\mathbf{V}^{\mathrm{T}}$. The first principal component explains the maximum amount of variance in the data. The second principal component explains the maximum amount of variance in the residual and so on. PCA, thus, yields a decomposition that is optimal in the least-mean-square sense. Therefore, it is often used for data reduction by keeping only the first l of the r principal components:

$$\mathbf{D}_l = \sum_{i=1}^{l} \mathbf{u}_i \sigma_i \mathbf{v}_i^{\mathrm{T}}\,.$$

In most cases, however, PCA provides no meaningful transformation of the data, not even after orthogonal rotation, for instance, of the basic waveforms that preserves their decorrelation but gives up the orthogonality of the topographies and the optimal decomposition in the least-mean-square sense. In contrast to PCA, ICA tries to find a meaningful decomposition by searching not only for decorrelated but for statistically independent waveforms. Statistical independence is a much stronger requirement than decorrelation as will be seen in the next section.

3.4.3. Statistical independence

The random variables s_1, s_2, ..., s_m are statistically independent if their joint probability density function $p(s_1,\ldots,s_m)$ can be factorized (Papoulis, 1991):

$$p(s_1,\ldots,s_m) = \prod_m p_i(s_i)$$

where $p_i(s_i)$ denotes the marginal density of s_i. Statistical independence is illustrated in the following example adapted from an ICA demo found at http://www.cis.hut.fi/projectes/ica/icademo. In the left panel of Fig. 3.3 two observed waveforms are depicted that are known to be the linear mixture of two statistically independent signals. The mixture is statistically dependent as each waveform carries information that is also contained in the other waveform. In the right panel the two-dimensional joint pdf of the signals is shown generated by plotting the amplitude of the bottom signal (x axis) versus the amplitude of the top signal (y axis) at each time point. Below the x axis and to the left of the y axis the marginal densities of the bottom and top signal are depicted. They show the frequency of occurrence for each amplitude value.

In Fig. 3.4 the waveforms of the previous figure are displayed after decorrelation together with their joint and marginal densities. Despite decorrelation each waveform provides a considerable amount of information about the other waveform. Consequently, the waveforms are still statistically dependent.

Finally, in Fig. 3.5 the original independent signals are shown. The waveforms, an impulsive noise signal and a sinusoid, are clearly separated. According to the definition of statistical independence the joint density can now be seen to be the product of the marginal densities in contrast to the two preceding figures.

Comparing the marginal densities in Fig. 3.3 and Fig. 3.5 reveals that the linear mixtures are more Gaussian than the original independent signals. The same phenomenon is expressed by the central limit theorem. Thus, in order to derive the independent components ICA has to maximize the non-normality of the marginal densities. This also explains why at most one independent signal underlying the mixture may be normally distributed (cf. chapter 3.4.1). Moreover, the example clearly illustrates that statistical independence is a much stronger requirement than uncorrelatedness. Only for normally distributed random variables that are completely defined by their first and second moment as higher moments are zero, decorrelation, i.e. a zero second moment, implies statistical independence. For any other distribution there can still be statistical dependence in spite of decorrelation (cf. Fig. 3.4).

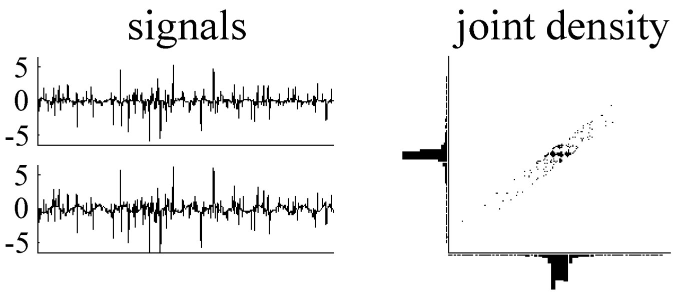

Fig. 3.3: Observed signals and density. (Left) Two observed signals generated by linear mixture of two statistically independent waveforms. (Right) The two-dimensional joint density of the observed signals and their respective marginal densities.

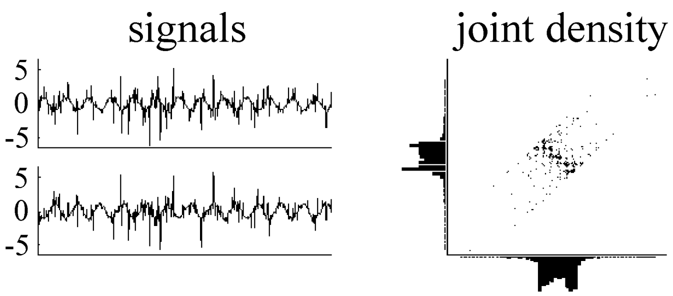

Fig. 3.4: Decorrelated signals and density. (Left) The two signals of Fig. 3.3 after decorrelation. (Right) Their two-dimensional joint density and marginal densities.

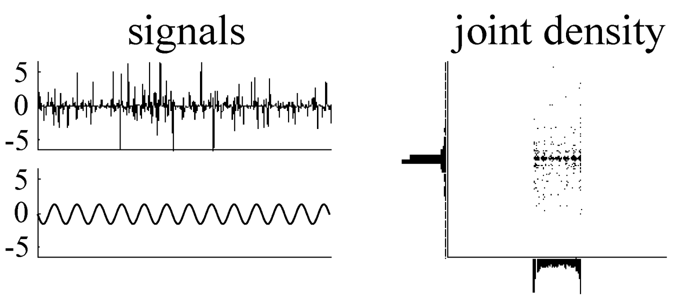

Fig. 3.5: Independent signals and density. (Left) The independent signals underlying the observed signals of Fig. 3.3. (Right) Their two-dimensional joint density and marginal densities.

3.4.4. Contrast functions

ICA is performed by optimizing a suitable contrast function. The contrast function is designed in such a way that the independent components are achieved when the function reaches its optimum. There are a couple of different contrast functions based, for instance, on maximum likelihood (e.g. Pham et al., 1992), the infomax principle (e.g. Nadal and Parga, 1994; Bell and Sejnowski, 1995), or higher order approximation of mutual information by cumulants (e.g. Comon, 1992, 1994; Cardoso and Souloumiac, 1993). In the next sections these approaches are introduced. It is pointed out that maximum likelihood, infomax and mutual information are equivalent under certain conditions. The presented maximum likelihood contrast ϕ_{L} and the cumulant-based contrast $\phi_{\mathrm{ICA}}^{\mathrm{o}}$ will appear again in chapter 6 in connection with the new SCICA algorithm.

Maximum likelihood

In the maximum likelihood approach it is assumed that the *m* (marginal) pdfs $\mathrm{p}_i(.)$ of the unknown independent waveforms are a priori known. With the knowledge of the pdfs it is possible to specify for any decomposition of the observed signals $\mathbf{D} = \mathbf{C}\mathbf{S}$ with $\mathbf{C} \in \Re^{m \times m}$ and $\mathbf{S} \in \Re^{m \times t}$ the joint pdf q(.) of the observed signals (Papoulis, 1991):

$$\mathrm{q}(\mathbf{d}_j) = \left|\det \mathbf{W}\right| \prod_{i=1}^{m} \mathrm{p}_i(\mathbf{w}_i^{\mathrm{T}} \mathbf{d}_j)$$

denoting by $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_m)^{\mathrm{T}}$ the matrix $\mathbf{C}^{-1}$ and by $\mathbf{d}_j$ the j^{th} column vector of **D**, i.e. the observed vector over all channels at time point *j*. The likelihood L, finally, is the product of the joint density q(.) over all samples

$$\mathrm{L} = \prod_{j=1}^{t} \mathrm{q}(\mathbf{d}_j) .$$

In Pham et al. (1992) the normalized (i.e. divided by *t* samples) log-likelihood whose gradient can be easily derived for optimization was introduced as a contrast function ϕ_{L}:

$$\phi_{\mathrm{L}} = \frac{1}{t} \ln \mathrm{L} = \frac{1}{t} \sum_{j=1}^{t} \sum_{i=1}^{m} \ln \mathrm{p}_i\left(\mathbf{w}_i^{\mathrm{T}} \mathbf{d}_j\right) + \ln\left|\det(\mathbf{W})\right| = \max .$$

The likelihood can be understood as a measure of the probability of a given decomposition. Maximizing the likelihood with respect to **W** using a suitable assumption about the pdfs of the unknown independent waveforms yields the required independent components. The pdfs need not be determined with great precision. In fact it is sufficient to estimate whether the pdf of the i^{th} independent component is super- or sub-Gaussian, i.e. has a positive or negative kurtosis (for definition of kurtosis refer to Appendix A1 in 3.4.7), and then to use a given super- or sub-Gaussian pdf instead of the real pdf (Girolami, 1998; Lee et al., 1999).

Infomax

The infomax approach has been derived from the neural network viewpoint (Nadal and Parga, 1994; Bell and Sejnowski, 1995). It is based on maximizing the output entropy (or information flow) of a single-layer feedforward neural network with non-linear units. The *m* observed signals $\mathbf{D} \in \Re^{m \times t}$ are the input to the neural network. The m output units transform the input according to $\mathrm{g}_i(\mathbf{w}_i^{\mathrm{T}}\mathbf{D})$ where $\mathrm{g}_i(.)$ is a non-linear function and $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_m)^{\mathrm{T}} \in \Re^{m \times m}$ is the unmixing matrix. The infomax contrast function ϕ_{I} can be summarized in

$$\phi_{\mathrm{I}} = \mathrm{H}(\mathrm{g}_1(\mathbf{w}_1^{\mathrm{T}}\mathbf{D}), \dots, \mathrm{g}_m(\mathbf{w}_m^{\mathrm{T}}\mathbf{D})) = \max$$

with differential entropy H. If the g_i are well chosen, maximizing H with respect to **W** minimizes the statistical dependence of the output waveforms. Several authors, e.g. Cardoso (1997), have proven that the infomax principle and the maximum likelihood approach are equivalent if the non-linearities $\mathrm{g}_i(.)$ are chosen as the cumulative distribution function of the pdfs $\mathrm{p}_i(.)$ used in the maximum likelihood estimate:

$$\mathrm{g}_i'(.) = \mathrm{p}_i(.)\,.$$

In the original infomax approach of Bell and Sejnowski (1995) only non-linearities corresponding to super-Gaussian pdfs were used. Original infomax is therefore not able to recover sub-Gaussian signals such as line noise in EEG recordings, for instance. The extended infomax approach (Girolami, 1998; Lee et al., 1999) overcomes this limitation by switching as required between a super- and sub-Gaussian pdf. The decision whether a pdf is modeled as sub- or super-Gaussian is taken in each optimization step, for instance by using a switching rule (Lee et al., 1999) that guarantees local stability of the unmixing matrix **W** (Cardoso and Laheld, 1996; Cardoso, 1998). The switching rule will be detailed as part of the extended infomax algorithm in chapter 3.4.5.

Cumulant-based contrasts derived from higher order approximation of mutual information

Mutual information is the natural measure of independence. It is always non-negative, and zero if and only if statistical independence is achieved. Mutual information is defined as the Kullback-Leibler divergence between a distribution and the closest distribution with independent entries. Cardoso (1998) has shown that the principles of mutual information and maximum likelihood are essentially equivalent provided that the pdfs in the maximum likelihood approach are accurately determined. As mutual information itself is difficult to estimate some authors have approximated it by polynomial density expansions (e.g. Comon, 1992, 1994) leading to higher order approximations of mutual information by cumulants (for definition of cumulants, see Appendix A2 in 3.4.7). Simplifying such a higher-order approximation, Comon (1994) has derived a fourth-order cumulant-based contrast function. This contrast does not necessarily approximate mutual information but minimizes statistical dependence of the m waveforms $\mathbf{S} = \mathbf{W}\,\mathbf{D}$ with unmixing matrix $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_m)^{\mathrm{T}} \in \Re^{m \times m}$ and observed data matrix $\mathbf{D} \in \Re^{m \times t}$. Statistical dependence is minimized by either minimizing the sum of the squared fourth-order cross-cumulants of the decomposed waveforms:

$$\phi^{\mathrm{o}}_{\mathrm{ICA}} = \sum_{\substack{ijkl=1 \\ ijkl \neq iiii}}^{m} \mathrm{Cum}^2\left[\mathbf{w}_i^{\mathrm{T}}\mathbf{D}, \mathbf{w}_j^{\mathrm{T}}\mathbf{D}, \mathbf{w}_k^{\mathrm{T}}\mathbf{D}, \mathbf{w}_l^{\mathrm{T}}\mathbf{D}\right] = \min$$

or equivalently by maximizing the sum of the smaller set of squared fourth-order auto-cumulants with respect to $\mathbf{W}$:

$$\phi^{\mathrm{o}}_{\mathrm{ICA}} = \sum_{i=1}^{m} \mathrm{Cum}^2\left[\mathbf{w}_i^{\mathrm{T}}\mathbf{D}, \mathbf{w}_i^{\mathrm{T}}\mathbf{D}, \mathbf{w}_i^{\mathrm{T}}\mathbf{D}, \mathbf{w}_i^{\mathrm{T}}\mathbf{D}\right] = \max .$$

Both contrast functions are orthogonal (as denoted by °) meaning that they are only valid under the whiteness constraint $\mathrm{E}\left\{\mathbf{S}\,\mathbf{S}^{\mathrm{T}}\right\} = \mathbf{I}$. Cardoso and Souloumiac (1993) have introduced another orthogonal contrast function operating on a subset of fourth-order cumulants:

$$\phi^{\mathrm{o}}_{\mathrm{JADE}} = \sum_{\substack{ijkl=1 \\ ijkl \neq ijkk}}^{m} \mathrm{Cum}^2\left[\mathbf{w}_i^{\mathrm{T}}\mathbf{D}, \mathbf{w}_j^{\mathrm{T}}\mathbf{D}, \mathbf{w}_k^{\mathrm{T}}\mathbf{D}, \mathbf{w}_l^{\mathrm{T}}\mathbf{D}\right] = \min$$

respectively

$$\phi^{\mathrm{o}}_{\mathrm{JADE}} = \sum_{ijk=1}^{m} \mathrm{Cum}^2\left[\mathbf{w}_i^{\mathrm{T}}\mathbf{D}, \mathbf{w}_j^{\mathrm{T}}\mathbf{D}, \mathbf{w}_k^{\mathrm{T}}\mathbf{D}, \mathbf{w}_k^{\mathrm{T}}\mathbf{D}\right] = \max .$$

This particular subset enables the usage of a specific algorithm called JADE that will be described in the next chapter.

The cumulant-based contrasts differ from other contrasts derived from maximum likelihood or infomax, for instance, inasmuch as they require no assumption about the pdf of the independent components but explicitly estimate the higher-order statistics from the decomposed data by means of cumulants.

3.4.5. ICA optimization algorithms

After choosing an ICA contrast function, a practicable optimization algorithm is needed. The optimization problem is treated in this chapter. As most ICA algorithms require or perform some preprocessing before optimization, preprocessing of the observed data is discussed in the first section. In the second section algorithms for maximum likelihood or infomax estimation are presented focusing on the extended infomax algorithm (Girolami, 1998; Lee et al., 1999) and a switching rule used therein (Lee et al. 1999) that guarantees local stability of the unmixing matrix (Cardoso and Laheld, 1996; Cardoso, 1998). The switching rule is also applied in the new SCICA algorithm (cf. chapter 6). The last section is concerned with optimization of fourth-order cumulant-based algorithms concentrating mainly on the joint diagonalization of cumulant matrices as performed in the JADE algorithm (Cardoso and Souloumiac, 1993). There are a couple of further ICA contrast functions and corresponding optimization algorithms reviewed e.g. by Hyvärinen (1999) that are not treated here as they are of no concern for the present work.

Preprocessing of the data

As has already been mentioned in the initial chapter 3.4.1 the basic preprocessing step of ICA is to center the observed data waveforms. In most ICA algorithms the centered data $\mathbf{D} \in \Re^{m \times t}$ are additionally whitened (sphered) before optimization

$$\overline{\mathbf{D}} = \overline{\mathbf{W}}\mathbf{D}$$

where $\overline{\mathbf{D}} \in \Re^{m \times t}$ are the whitened signals with $\mathrm{E}\{\overline{\mathbf{D}}\,\overline{\mathbf{D}}^{\mathrm{T}}\} = \mathbf{I}$. Centered and whitened waveforms have a unit variance and are decorrelated. The whitening matrix $\overline{\mathbf{W}} \in \Re^{m \times m}$ may be obtained from the SVD $\frac{1}{t}\mathbf{D}\mathbf{D}^{\mathrm{T}} = \mathbf{U}\Sigma\mathbf{U}^{\mathrm{T}}$ setting $\overline{\mathbf{W}} = \Sigma^{-\frac{1}{2}}\mathbf{U}^{\mathrm{T}}$. Whitening $\mathbf{D}$ changes the ICA problem to

$$\mathbf{S} = \mathbf{W}\,\overline{\mathbf{D}} = \mathbf{W}\,\overline{\mathbf{W}}\,\mathbf{D}.$$

Only if $\mathbf{W} \in \Re^{m \times m}$ is a rotation (or orthogonal) matrix, i.e. $\mathbf{W}\mathbf{W}^{\mathrm{T}} = \mathbf{W}^{\mathrm{T}}\mathbf{W} = \mathbf{I}$, it is guaranteed that $\mathrm{E}\{\mathbf{S}\mathbf{S}^{\mathrm{T}}\} = \mathbf{I}$ which is a prerequisite for orthogonal contrasts such as $\phi^{\mathrm{o}}_{\mathrm{ICA}}$ or $\phi^{\mathrm{o}}_{\mathrm{JADE}}$. Thus, under the whiteness constraint $\mathrm{E}\{\mathbf{S}\mathbf{S}^{\mathrm{T}}\} = \mathbf{I}$ the ICA problem reduces to whitening the data and finding a suitable orthogonal matrix **W**. For non-orthogonal contrasts whitening is, in principle, not necessary but improves convergence of the optimization algorithm.

Sometimes it may also be useful to reduce the dimension of the centered data **D** in combination with whitening by truncating the SVD (PCA) of **D** at component $l <$ rank(**D**)

$$\mathbf{D}_l = \sum_{i=1}^{l} \mathbf{u}_i \sigma_i \mathbf{v}_i^{\mathrm{T}}$$

and setting $\overline{\mathbf{D}} = \sqrt{t}\,(\mathbf{v}_1, \ldots, \mathbf{v}_l)^{\mathrm{T}} \in \Re^{l \times t}$ which is equivalent to replacing the m observed waveforms by their first l whitened basic waveforms (Comon, 1994). In the remainder of chapter 3.4 especially when talking about the fourth-order cumulant-based algorithms it is assumed that the data has been preprocessed by centering and whitening. For simplicity of notation, the preprocessed data are denoted again by **D**.

Algorithms for maximum likelihood or infomax estimation

Algorithms for maximizing the log-likelihood or infomax contrasts ϕ_{L} and ϕ_{I} with respect to **W** are usually based on gradient ascent of the contrast function. An update rule for **W** based on stochastic gradient ascent was derived e.g. in Bell and Sejnowski (1995):

$$\Delta\mathbf{W} \propto \left(\mathbf{W}^{\mathrm{T}}\right)^{-1} - \varphi(\mathbf{WD})\mathbf{D}^{\mathrm{T}}.$$

Despite whitening of the input data **D** this algorithm may converge very slowly. Convergence can be improved by following the natural (Amari et al., 1996; Amari, 1998) or relative gradient (Cardoso and Laheld, 1996) which gives the steepest direction of the target function. Both principles amount to multiplying the right-hand side of the previous learning rule by $\mathbf{W}^{\mathrm{T}}\mathbf{W}$ yielding:

$$\Delta\mathbf{W} \propto [\mathbf{I} - \varphi(\mathbf{WD})(\mathbf{WD})^{\mathrm{T}}]\,\mathbf{W}\ .$$

The entry-wise non-linear function $\varphi(\mathbf{S}) = [\varphi_1(\mathbf{s}_1),\ldots,\varphi_m(\mathbf{s}_m)]$ collects the score functions related to the m decomposed waveforms $\mathbf{S} = \mathbf{W}\,\mathbf{D}$ with $\mathbf{S} = (\mathbf{s}_1,\ldots,\mathbf{s}_m)^{\mathrm{T}} \in \Re^{m \times t}$. The score functions are defined as $\varphi_i(.) = -(\ln \mathrm{p}_i(.))'$ or equivalently $\varphi_i(.) = \dfrac{-\mathrm{p}_i'(.)}{\mathrm{p}_i(.)}$ where p_i is the pdf of the i^{th} decomposed waveform. The extended infomax algorithm (Girolami, 1998; Lee et al., 1999) switches between two score functions derived from a super- and sub-Gaussian pdf. A pair of score functions is, for example (Girolami, 1998; Lee et al., 1999):

$$\varphi_i(\mathbf{s}_i) = \mathbf{s}_i + u_i \tanh(\mathbf{s}_i) \quad \begin{cases} u_i = +1 & : \ \text{super} - \text{Gaussian} \\ u_i = -1 & : \ \text{sub} - \text{Gaussian} \end{cases}$$

corresponding to the super- and sub-Gaussian pdfs:

$$\mathrm{p}_{\text{super}}(\mathbf{s}) \propto \mathrm{N}(0,1)\,\mathrm{sech}^2(\mathbf{s}) \ \text{ and } \ \mathrm{p}_{\text{sub}}(\mathbf{s}) = \frac{1}{2}\left(\,\mathrm{N}(1,1) + \mathrm{N}(-1,1)\,\right)$$

where $\mathrm{N}(\mu,\sigma^2)$ is the Gaussian density with mean μ and variance σ^2. The switching rule proposed in Lee et al. (1999)

$$u_{\mathrm{i}} = \mathrm{sign}\!\left(\mathrm{E}\{\mathrm{sech}^2(\mathbf{s}_i)\}\,\mathrm{E}\{\mathbf{s}_i^2\} - \mathrm{E}\{\tanh(\mathbf{s}_i)\mathbf{s}_i\}\right)$$

is based on the stability analysis of Cardoso and Laheld (1996). Stability analysis deals with the local stability of the unmixing matrix. The unmixing matrix is considered to be locally stable if small perturbations from equilibrium are pulled back to the separating point. A sufficient stability condition is (Cardoso and Laheld, 1996; Cardoso, 1998)

$$\kappa_i = \mathrm{E}\left\{\varphi_i'(\mathbf{s}_i)\right\}\mathrm{E}\left\{\mathbf{s}_i^2\right\} - \mathrm{E}\left\{\varphi_i(\mathbf{s}_i)\mathbf{s}_i\right\} > 0$$

for all waveforms $\mathbf{s}_i$ with $i = 1\ldots m$. Substituting the score function into the stability condition gives (Lee et al., 1999)

$$\begin{aligned}\kappa_i &= \mathrm{E}\left\{1 + u_i \operatorname{sech}^2(\mathbf{s}_i)\right\}\mathrm{E}\left\{\mathbf{s}_i^2\right\} - \mathrm{E}\left\{\left[\mathbf{s}_i + u_i \tanh(\mathbf{s}_i)\right]\mathbf{s}_i\right\} \\ &= u_i\left(\mathrm{E}\left\{\operatorname{sech}^2(\mathbf{s}_i)\right\}\mathrm{E}\left\{\mathbf{s}_i^2\right\} - \mathrm{E}\left\{\tanh(\mathbf{s}_i)\,\mathbf{s}_i\right\}\right).\end{aligned}$$

To ensure $\kappa_i > 0$ the sign of $\mathrm{E}\left\{\operatorname{sech}^2(\mathbf{s}_i)\right\}\mathrm{E}\left\{\mathbf{s}_i^2\right\} - \mathrm{E}\left\{\tanh(\mathbf{s}_i)\,\mathbf{s}_i\right\}$ must be the same as the sign of u_i. This yields exactly the above switching rule.

A Matlab toolbox containing the infomax and extended infomax algorithm is available at http://www.cnl.salk.edu/~scott/ica.html. The implemented extended infomax algorithm uses a switching rule based on the sign of the kurtosis estimated for each $\mathbf{s}_i$. This switching rule, however, does not necessarily guarantee stability of the unmixing matrix (Cardoso, 1998).

Fourth-order cumulant-based algorithms

The orthogonal fourth-order cumulant-based contrast functions $\phi^{\mathrm{o}}_{\mathrm{ICA}}$ and $\phi^{\mathrm{o}}_{\mathrm{JADE}}$ allow for the usage of special optimization techniques. Maximizing the sum of all squared fourth-order auto-cumulants as in $\phi^{\mathrm{o}}_{\mathrm{ICA}}$ may, for instance, be reduced to a pairwise optimization problem (Comon, 1994). Starting from an initial decomposition of the data $\mathbf{D} = \mathbf{C}\mathbf{S}$ with $\mathbf{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_m) \in \Re^{m\times m}$ and $\mathbf{S} \in \Re^{m\times t}$ each pair of topographies $(\mathbf{c}_i, \mathbf{c}_j)$ with $j > i$ is rotated by a suitable Givens rotation such that the sum of fourth-order auto-cumulants of both corresponding (rotated) waveforms is maximal with the optimal angle at each step being available in closed form. This sequence of pairwise rotations is repeated until the rotation angles are negligibly small. The particular subset of fourth-order cumulants in the JADE contrast $\phi^{\mathrm{o}}_{\mathrm{JADE}}$ may be optimized by joint diagonalization of m^2 cumulant matrices $\mathbf{N}_{kl} \in \Re^{m\times m}$:

$$\phi^{\mathrm{o}}_{\mathrm{JADE}} = \sum_{ijk=1}^{m} \mathrm{Cum}^2\left[\mathbf{s}_i, \mathbf{s}_j, \mathbf{s}_k, \mathbf{s}_k\right] = \sum_{kl=1}^{m} \left\|\operatorname{diag}(\mathbf{W}^{\mathrm{T}}\mathbf{N}_{kl}\mathbf{W})\right\| = \max$$

where diag(.) is the vector built from the diagonal of the matrix argument. The joint diagonalizer $\mathbf{W} \in \Re^{m\times m}$ is equal to the unmixing matrix with $\mathbf{S} = \mathbf{W}\,\mathbf{D}$ and $\mathbf{S} = \left(\mathbf{s}_1, \ldots, \mathbf{s}_m\right)^{\mathrm{T}} \in \Re^{m\times t}$. The cumulant matrices $\mathbf{N}_{kl} = (n_{ij}^{(kl)})$ are derived according to $n_{ij}^{(kl)} = \mathrm{Cum}[\mathbf{s}_i, \mathbf{s}_j, \mathbf{s}_k, \mathbf{s}_l]$ for $ijkl = 1\ldots m$. Exploiting the symmetries $\mathbf{N}_{kl} = \mathbf{N}_{lk}$ of the cumulant matrices (cf. Appendix A2 in 3.4.7) it is sufficient to diagonalize the subset of $m(m+1)/2$ cumulant matrices with $l \geq k$:

$$\sum_{\substack{kl=1 \\ l \geq k}}^{m} \left\| \operatorname{diag}(\mathbf{W}^{\mathrm{T}} \mathbf{N}_{kl} \mathbf{W}) \right\| = \max .$$

Diagonalization may be performed by successive pairwise Givens rotations of the columns of $\mathbf{W}$ with the optimal angle in each step being available in closed form. In the noise-free or low-noise case the dimension of the problem may be further reduced by approximating the m^2 cumulant matrices by their first m eigenmatrices and diagonalizing these m eigenmatrices (Cardoso and Souloumiac, 1993). Hence the name JADE: **j**oint **a**pproximate **d**iagonalization of **e**igenmatrices. A Matlab implementation of the JADE algorithm based on diagonalization of $m(m+1)/2$ cumulant matrices is available at http://sig.enst.fr/~cardoso/stuff.html.

3.4.6. Artifact correction by ICA

Application of ICA to artifact correction has been initiated by Makeig et al. (1996) and Vigário (1997) who realized that brain and non-brain activity often separate into different independent components. Independent components have to be manually classified as artifact or brain signal components. Artifact correction is, finally, accomplished by multiplying the matrix of p brain signal topographies $\mathbf{B} \in \Re^{m \times p}$ with their corresponding independent waveforms $\mathbf{S}_{\mathrm{B}} \in \Re^{p \times t}$ (Jung et al., 1998, 2000a,b):

$$\mathbf{D}_{\mathrm{corr}} = \mathbf{B}\mathbf{S}_{\mathrm{B}} .$$

Note that this is just the opposite way to MSEC where the matrix of n artifact topographies $\mathbf{A} \in \Re^{m \times n}$ is multiplied with the corresponding artifact waveforms $\mathbf{S}_{\mathrm{A}} \in \Re^{n \times t}$ and is then subtracted from the original data (see 3.3.2). In principle, however, ICA artifact correction could also be performed the latter way.

ICA can be applied to the correction of all artifacts whose time courses are independent from cerebral activity irrespective of their spatial correlation. This includes also muscle artifacts and even sub-Gaussian line noise provided that the extended infomax algorithm is applied (Jung et al., 1998, 2000a). The main assumption for artifact and brain activity to separate into distinct components is their independence. In continuous recordings decomposing 10-s epochs artifact and cerebral activity are in general sufficiently independent (Jung et al., 2000a). Artifact and cerebral waveforms in averaged event-related recordings, however, may be highly dependent. ICA is therefore not suitable for artifact correction in averaged data. Alternatively, it can be applied to continuous event-related recordings before averaging (Jung et al., 2000b). The major drawback of artifact correction by ICA is the a posteriori visual identification of artifact components that can be very time-consuming (Jung et al., 2000a).

3.4.7. Appendix

A1: Kurtosis

The kurtosis of the pdf p(s) of a random variable s is equal to the normalized fourth-order auto-cumulant of p(s) (Kendall and Stuart, 1977):

$$\mathrm{K}[s] = \frac{\mathrm{Cum}[s,s,s,s]}{\mathrm{E}^2\left\{s^2\right\}}$$

where the brackets [.] used with K[.] and Cum[.] denote that kurtosis and cumulants are functions of the distribution of s. Kurtosis can be considered a measure of the non-Gaussianity of a pdf. For a Gaussian pdf kurtosis is zero. It is typically positive for unimodal symmetric pdfs that are more sharply peaked than the Gaussian pdf and negative for unimodal symmetric pdfs that are flatter around the mean than the Gaussian density. Pdfs of positive (respectively negative) kurtosis are called super-Gaussian (respectively sub-Gaussian). In Fig. 3.6 a sub- and super-Gaussian pdf are depicted together with the Gaussian density.

A2: Cumulants

Consider the vector $\mathbf{t} = \left(t_1, \dots, t_m\right)^{\mathrm{T}} \in \Re^m$ and the vector $\mathbf{s} = \left(s_1, \dots, s_m\right)^{\mathrm{T}} \in \Re^m$ of m zero-mean random variables s_{i} whose characteristic function is denoted by $\hat{\mathrm{f}}(t)$:

$$\hat{\mathrm{f}}(t) = \mathrm{E}\left\{\exp(\mathrm{i}\mathbf{t}^{\mathrm{T}}\mathbf{s})\right\}.$$

The k^{th} order cumulant $\mathrm{Cum}[s_{i_1}, \dots, s_{i_k}]$ of the random variables $s_1, \dots, s_m$ with $i_1, \dots, i_k = 1 \dots m$ is defined as the coefficient of the term $t_{i_1}, \dots, t_{i_k}$ in the Taylor series expansion of the cumulant-generating function k(t) (Kendall and Stuart, 1977):

$$\mathrm{k}(t) = \ln \hat{\mathrm{f}}(t) \ .$$

The first-, second-, third- and fourth-order cumulants, for example, are given by

$$\mathrm{Cum}[s_{i_1}] = \mathrm{E}\{s_{i_1}\}$$

$$\mathrm{Cum}[s_{i_1}, s_{i_2}] = \mathrm{E}\{s_{i_1} s_{i_2}\}$$

$$\mathrm{Cum}[s_{i_1}, s_{i_2}, s_{i_3}] = \mathrm{E}\{s_{i_1} s_{i_2} s_{i_3}\}$$

$$\mathrm{Cum}[s_{i_1}, s_{i_2}, s_{i_3}, s_{i_4}] = \mathrm{E}\{s_{i_1} s_{i_2} s_{i_3} s_{i_4}\} - \mathrm{E}\{s_{i_1} s_{i_2}\}\,\mathrm{E}\{s_{i_3} s_{i_4}\} - \mathrm{E}\{s_{i_1} s_{i_3}\}\,\mathrm{E}\{s_{i_2} s_{i_4}\} - \mathrm{E}\{s_{i_1} s_{i_4}\}\,\mathrm{E}\{s_{i_2} s_{i_3}\}\ .$$

The auto-cumulants of s_i of order k=1...4 are obtained by dropping the distinct indices:

$$\mathrm{Cum}[s_i] = \mathrm{E}\{s_i\}$$

$$\mathrm{Cum}[s_i, s_i] = \mathrm{E}\{s_i^2\}$$

$$\mathrm{Cum}[s_i, s_i, s_i] = \mathrm{E}\{s_i^3\}$$

$$\mathrm{Cum}[s_i, s_i, s_i, s_i] = \mathrm{E}\left\{s_i^4\right\} - 3\mathrm{E}^2\left\{s_i^2\right\}.$$

Any cumulant of at least two random variables is called a cross-cumulant. Note that auto-cumulants up to order k=3 are equal to the k^{th} moment of s_i defined as $\mathrm{E}\left\{s_i^k\right\}$, i.e. the mean ($k$=1), variance ($k$=2) and skewness ($k$=3) of the distribution of s_i. The normalized, i.e. divided by the squared variance, fourth-order auto-cumulant is also referred to as kurtosis (cf. A1).

In the case of non zero-mean random variables s_i, s_{i_k} in the above formulae has to be replaced by $s_{i_k} - \mathrm{E}\{s_{i_k}\}$ yielding, for instance, for the fourth-order auto-cumulant:

$$\mathrm{Cum}[s_i, s_i, s_i, s_i] = \mathrm{E}\left\{s_i^4\right\} - 3\mathrm{E}^2\left\{s_i^2\right\} + 12\mathrm{E}^2\left\{s_i\right\}\mathrm{E}\left\{s_i^2\right\} - 4\mathrm{E}\left\{s_i\right\}\mathrm{E}\left\{s_i^3\right\} + 6\mathrm{E}^4\left\{s_i\right\}.$$

The cumulants are symmetric in their arguments, i.e. $\mathrm{Cum}[s_1, \ldots, s_k] = \mathrm{Cum}[s_{j_1}, \ldots, s_{j_k}]$ where $(j_1, \ldots, j_k)$ is a permutation of $(1, \ldots, k)$. This property is exploited in the JADE algorithm (chapter 3.4.5) in order to reduce the number of cumulant matrices from m^2 to $m(m+1)/2$. Further properties of cumulants may be found in Kendall and Stuart (1977).

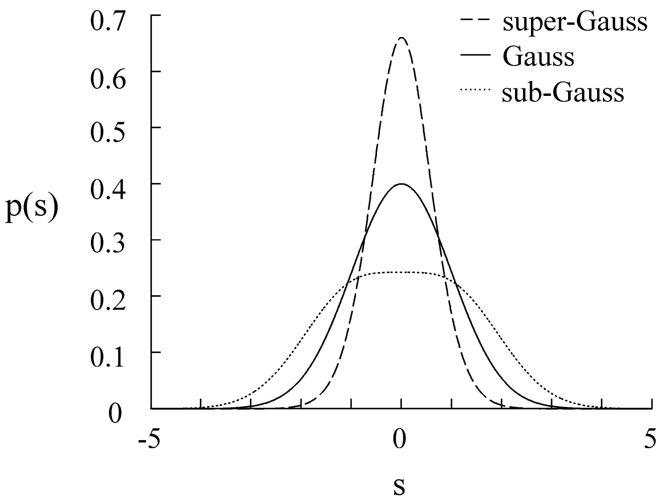

Fig. 3.6: The super-Gaussian density function (kurtosis>0) is more sharply peaked than the Gaussian density function (kurtosis=0). The sub-Gaussian density (kurtosis<0) is flatter around the mean than the Gaussian density.

4. Spatial filter for artifact correction

In chapter 4 the spatial filter for artifact correction used in this thesis is described in detail starting in section 4.1 with the basic model underlying artifact correction. The spatial filter is based on artifact and brain signal topographies to enable a distortion-free separation of artifact and cerebral activity. It differs from earlier approaches such as MSEC or ICA that have been described in the previous chapter in the way artifact and brain signal topographies are estimated. Comparable to MSEC, artifact topographies are derived from single or averaged artifact prototypes. The calculation of artifact topographies is described in section 4.2. In order to estimate brain signal topographies two new approaches are introduced in this thesis: the preselection approach and SCICA. Section 4.3 points out their basic ideas. The new approaches are described in detail in chapters 5 and 6. The spatial filter operator itself is introduced in section 4.4 in compact mathematical formulation. The same operator may be used with artifact and brain signal topographies determined by MSEC and ICA.

4.1. Basic model

The artifact-contaminated activity $\mathbf{D} \in \Re^{m \times t}$ measured in the EEG or MEG at m electrodes/sensors and t time points is assumed to be the weighted sum of n^* unknown artifact signals $\mathbf{S}_{\mathrm{A}*} = \left(\mathbf{s}_{\mathrm{A}*,1}, \ldots, \mathbf{s}_{\mathrm{A}*,n*}\right)^{\mathrm{T}} \in \Re^{n^* \times t}$ and p^* unknown brain signals $\mathbf{S}_{\mathrm{B}*} = \left(\mathbf{s}_{\mathrm{B}*,1}, \ldots, \mathbf{s}_{\mathrm{B}*,p*}\right)^{\mathrm{T}} \in \Re^{p^* \times t}$:

$$\mathbf{D} = \mathbf{A}^* \mathbf{S}_{\mathrm{A}*} + \mathbf{B}^* \mathbf{S}_{\mathrm{B}*}$$

where $\mathbf{A}^* \in \Re^{m \times n^*}$ and $\mathbf{B}^* \in \Re^{m \times p^*}$ are the unknown mixing matrices. The i^{th} column vector of $\mathbf{A}^*$ ($\mathbf{B}^*$) describes the topography of the i^{th} waveform $\mathbf{s}_{\mathrm{A}*,i}$ ($\mathbf{s}_{\mathrm{B}*,i}$). Contributions to $\mathbf{D}$ of artifact and cerebral origin can be separated by estimating two sets of topographies $\mathbf{A} \in \Re^{m \times n}$ and $\mathbf{B} \in \Re^{m \times p}$ with $\operatorname{rank}(\mathbf{A}) = n$, $\operatorname{rank}(\mathbf{B}) = p$ and $n + p \leq m$ such that $\operatorname{range}(\mathbf{A})$ is an estimate of $\operatorname{range}(\mathbf{A}^*)$ and $\operatorname{range}(\mathbf{B})$ is an estimate of $\operatorname{range}(\mathbf{B}^*)$.

First the artifact subspace $\operatorname{range}(\mathbf{A})$ is modeled using prior knowledge about artifacts that is not necessarily derived from $\mathbf{D}$. Each modeled artifact topography belongs to a basic artifact category. The spatial distribution of the artifacts is assumed to be nearly constant throughout the recording. Based on the predefined artifact subspace the brain signal subspace $\operatorname{range}(\mathbf{B})$ of the current data segment $\mathbf{D}$ is estimated. For that purpose it is further assumed in SCICA that the waveforms $\mathbf{S}_{\mathrm{A}} = (\mathbf{s}_{\mathrm{A},1},\ldots,\mathbf{s}_{\mathrm{A},n})^{\mathrm{T}} \in \Re^{n\times t}$ and $\mathbf{S}_{\mathrm{B}} = (\mathbf{s}_{\mathrm{B},1},\ldots,\mathbf{s}_{\mathrm{B},p})^{\mathrm{T}} \in \Re^{p\times t}$ corresponding to **A** and **B** with

$$\begin{pmatrix} \mathbf{S}_{\mathrm{A}} \\ \mathbf{S}_{\mathrm{B}} \end{pmatrix} = (\mathbf{A}\ \mathbf{B})^{\dagger}\ \mathbf{D}$$

are maximally independent where $(\mathbf{A}\ \mathbf{B})^{\dagger}$ denotes the Moore-Penrose pseudo-inverse of the compound matrix $(\mathbf{A}\ \mathbf{B}) \in \Re^{m\times(n+p)}$. Comparable to ICA (cf. chapter 3.4.1) it is assessed that each of the t samples of the decomposed signals $\mathbf{s}_{\mathrm{A},i}$ and $\mathbf{s}_{\mathrm{B},i}$ is an independent realization of the random variable $s_{\mathrm{A},i}$, $s_{\mathrm{B},i}$ distributed with pdf $\mathrm{p}_{\mathrm{A},i}(s_{\mathrm{A},i})$, $\mathrm{p}_{\mathrm{B},i}(s_{\mathrm{B},i})$.

4.2. Modeling the artifact subspace

Modeling the artifact subspace may be accomplished by calculating the eigenvectors (PCA topographies) of the moment matrix of suitable artifact prototypes. For eye artifacts, eigenvectors can be determined on individual calibration data sets containing blinks, horizontal and vertical eye movements. Oblique eye movements do not have to be modeled explicitly as they are inherently expressed as linear combinations of horizontal and vertical eye movements. For cardiac artifacts, eigenvectors are best derived from the average of several hundreds of cardiac cycles having an enhanced SNR compared to a single cardiac artifact. In general it is sufficient to describe ocular artifacts by the eigenvector with the maximum eigenvalue. Cardiac artifacts may have to be modeled by more than one eigenvector if further eigenvalues still explain a significant amount of the total variance. The n usually non-orthogonal artifact topographies (eigenvectors) are summarized in

$$\mathbf{A} = (\mathbf{a}_1,\ldots,\mathbf{a}_n) \in \Re^{m\times n}.$$

Linear independence of the $\mathbf{a}_i$, i.e. $\mathrm{rank}(\mathbf{A}) = n$, is guaranteed by adding only artifact topographies to **A** that cannot be expressed as a linear combination of the existing column vectors of **A**. The artifact subspace has to be determined only once for a recording session. Note that $\mathrm{range}(\mathbf{A}^{*})$ may be embedded into $\mathrm{range}(\mathbf{A})$, for instance if a data segment contains an oblique eye movement that is modeled as linear combination of a horizontal and vertical eye movement topography or if **A** encloses topographies that are not part of $\mathbf{A}^{*}$.

4.3. Modeling the signal subspace

The brain signal subspace is determined separately for each data segment to be corrected. In order to model the brain signal subspace the new preselection approach and the novel concept of SCICA are introduced.

In the preselection approach, modeling the brain signal subspace consists of extracting a relevant number p of eigenvectors (PCA topographies) from an artifact-free subset of the data segment. The subset is obtained by excluding sample vectors from the original data segment that exceed a certain amplitude or correlation with the artifact subspace. The preselection approach is described in detail in chapter 5.

In the SCICA approach, the brain signal subspace is estimated from the whole data segment. SCICA utilizes the prior knowledge about artifact topographies and combines this with the temporal-statistical strategy of ICA to estimate brain signal topographies. Starting from the n predefined artifact topographies, the artifact-contaminated data segment is decomposed iteratively into p further components until all waveforms are maximally independent under the spatial constraint. The additionally obtained components represent an estimate of cerebral activity. The new SCICA decomposition is described in detail in chapter 6.

The p linearly independent brain signal topographies are summarized in

$$\mathbf{B} = \left(\mathbf{b}_1, \ldots, \mathbf{b}_p\right) \in \Re^{m \times p}.$$

In the preselection approach the $\mathbf{b}_i$ are orthonormal.

4.4. Spatial filter

Once n artifact and p signal topographies are determined, a matrix $\mathbf{C}$ of both sets of topographies is composed:

$$\mathbf{C} = \left(\mathbf{A}\ \mathbf{B}\right) = \left(\mathbf{a}_1, \ldots, \mathbf{a}_n, \mathbf{b}_1, \ldots, \mathbf{b}_p\right) \in \Re^{m \times l}$$

with $n + p = l$ and inverted calculating the Moore-Penrose pseudo-inverse

$$\mathbf{C}^{\dagger} = \left(\mathbf{a}_1', \ldots, \mathbf{a}_n', \mathbf{b}_1', \ldots, \mathbf{b}_p'\right)^{\mathrm{T}} \in \Re^{lxm}.$$

Premultiplying the spatial filter

$$\mathbf{F}_1 = \left(\mathbf{a}_1', \ldots, \mathbf{a}_n'\right)^{\mathrm{T}} \in \Re^{n \times m}$$

to the EEG/MEG data matrix $\mathbf{D}$ yields the waveforms corresponding to the artifact topographies.

The artifact-free data segment is determined by premultiplying the spatial filter

$$\mathbf{F}_2 = \mathbf{I} - (\mathbf{a}_1, \ldots, \mathbf{a}_n)(\mathbf{a}_1', \ldots, \mathbf{a}_n')^{\mathrm{T}} \in \Re^{m \times m}$$

to the EEG/MEG data matrix **D** with identity matrix $\mathbf{I} \in \Re^{m \times m}$. Applying the spatial filter $\mathbf{F}_2$ is equivalent to subtracting the reconstructed artifact activity at the electrodes/sensors from the artifact-contaminated data **D**. In principle, it is also possible to reconstruct the artifact-free brain activity directly at each electrode/sensor using the spatial filter

$$\mathbf{F}_3 = (\mathbf{b}_1, \ldots, \mathbf{b}_p)(\mathbf{b}_1', \ldots, \mathbf{b}_p')^{\mathrm{T}} \in \Re^{m \times m}.$$

This approach employed e.g. with ICA (Jung et al., 1998, 2000a,b) is only advisable if brain activity is completely represented by the brain signal topographies. In the preselection approach, for instance, which involves reduction of the data to a subset and further approximation of the subset by p relevant eigenvectors, brain activity may be sufficiently modeled to guarantee distortion-free separation from artifact activity but not exhaustively to justify application of $\mathbf{F}_3$. The same reasoning applies if the dimension of data matrix **D** is reduced before estimating the brain signal topographies using truncated SVD, for instance. In these cases it is preferable to subtract the reconstructed artifact activity from the original data as it is modeled more precisely. In the rest of this thesis, therefore, only spatial filter $\mathbf{F}_2$ is applied.

5. A new approach: Modeling the signal subspace by preselection

In this chapter the novel preselection approach to model the brain signal subspace is introduced. The approach and its parameters are described in chapter 5.1. In chapter 5.2 two examples for spatial filtering with preselection are given. The examples illustrate the good performance of the method but also demonstrate its dependence on a suitable choice of the parameter thresholds. In chapter 5.3 the approach is discussed. The spatial filter with preselection has been published as the spatial component method in Ille et al. (1997).

5.1. Method

In the preselection approach, a relevant number of eigenvectors (PCA topographies) is determined from the moment matrix of an artifact-free subset of the data segment. The subset $\mathbf{D}_{\mathrm{sub}}$ is obtained by excluding sample vectors $\mathbf{d}_i \in \Re^m$ from the original data segment $\mathbf{D} = (\mathbf{d}_1, \ldots, \mathbf{d}_t) \in \Re^{m \times t}$ that exceed a certain amplitude at one or more electrodes/sensors or a specific correlation with the matrix of artifact topographies $\mathbf{A} \in \Re^{m \times n}$ defined in chapter 4.2. The amplitude and correlation criterion and the number of eigenvectors are discussed in the following sections.

5.1.1. Amplitude criterion

By means of the amplitude criterion high-amplitude artifacts exceeding the normal amplitude range of the EEG/MEG are identified, e.g. eye movements or blinks in the EEG or cardiac activity in the MEG. The amplitude threshold should be set to a value slightly above the maximum brain signal amplitude. In an adult's EEG the threshold is in general about 100 µV. In children's EEG a threshold may be more difficult to determine as signal amplitudes are usually higher and may interfere with the amplitude range of high-amplitude artifacts. In an MEG system with planar gradiometers the amplitude threshold usually lies between 500 and 800 fT/cm. The indicated thresholds hold for a filter setting of 1-70 Hz.

5.1.2. Correlation criterion

Sample vectors of artifact origin that have not been identified by the amplitude criterion may be detected by the correlation criterion. The (subspace) correlation can be calculated as the cosine of the principal angle between the sample vector $\mathbf{d}_i$ and the matrix of predefined artifact topographies $\mathbf{A}$ (Golub and van Loan, 1996). Alternatively, it may be derived from the scalar products of the normalized sample vector and the n vectors $\mathbf{a}_j^{\mathrm{o}}$ that form an orthonormal basis of range($\mathbf{A}$) according to

$$\mathrm{corr}_i = \sqrt{\sum_{j=1}^{n}\left(\frac{\mathbf{d}_i^{\mathrm{T}}}{\|\mathbf{d}_i\|}\mathbf{a}_j^{\mathrm{o}}\right)^2} \quad .$$

The $\mathbf{a}_j^{\mathrm{o}}$ can be determined by QR-decomposition of $\mathbf{A}$ (Golub and van Loan, 1996). The particular correlation threshold depends on the unknown correlation between cerebral and artifact activity and has to be determined empirically. The higher the correlation, the higher the threshold has to be set to allow the correlated sample vectors of cerebral origin to enter $\mathbf{D}_{\mathrm{sub}}$. With an adequate choice of the correlation threshold correlated brain activity remains undistorted by spatial filtering.

5.1.3. Number of eigenvectors

The number of eigenvectors that can be derived from the moment matrix of $\mathbf{D}_{\mathrm{sub}}$ is equal to rank($\mathbf{D}_{\mathrm{sub}}$). The eigenvectors mainly model cerebral activity but also noise and to some extent artifact activity that has not been detected by the amplitude and correlation criterion. Noise and remaining artifact activity are primarily represented by eigenvectors with low eigenvalues. Therefore only the first p of the eigenvectors sorted from highest to lowest eigenvalue are used as an estimate of the brain signal subspace in the spatial filter design. The model of the brain signal subspace containing p orthonormal eigenvectors can be summarized in

$$\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_p) \in \Re^{m \times p} \; .$$

The particular number p of eigenvectors depends on the unknown number of distinguishable cerebral sources and the unknown amount of artifact in $\mathbf{D}_{\mathrm{sub}}$. It has to be determined empirically. If too few eigenvectors are used, brain activity may be distorted. Applying too many eigenvectors may result in noise enhancement as will be demonstrated in the following chapter.

5.2. Examples

The spatial filter described in chapter 4 using preselection to model the brain signal subspace was implemented in C (Microsoft Visual C++ 6.0). In this section it is applied to two segments taken from spontaneous EEG recordings. The examples demonstrate the influence of correlation threshold and number of eigenvectors on artifact correction. In Example 1 (5.2.1) an incorrect correlation threshold results in a severe distortion of brain activity. In Example 2 (5.2.2) an inadequate number of eigenvectors leads to a considerable noise enhancement. In both examples, the quality of artifact correction is evaluated by thorough visual comparison of the original and artifact-corrected EEG segment. Artifacts are considered to be completely removed if they are reduced to below visibility. Brain activity is regarded as undistorted if no apparent changes in brain signals can be observed, i.e. if no spurious activity is added and no relevant brain activity is increased or decreased in amplitude or even removed.

5.2.1. Example 1: Influence of correlation threshold

In Fig. 5.1 the influence of the correlation threshold on artifact correction is demonstrated. In the upper left panel the original EEG segment (5 s, 1–70 Hz, referenced to the average of F3 and F4) containing a blink in the 1st second followed by an oblique eye movement in the 2nd second is depicted. Horizontal and vertical eye movement topographies were defined on single prototypes of the respective artifact category. In this example it is not necessary to define an additional blink topography as the blink is adequately modeled by the vertical eye movement topography. The amplitude threshold was set to 120 µV because the rhythmic brain activity starting in the 3rd second exhibits a maximum amplitude of about 110 µV. The correlation threshold was varied between 10 % and 90 % in steps of 20 %. The number of eigenvectors was set to 3 giving the best results with a suitable correlation threshold. The subsequent panels of Fig. 5.1 show the data segment after artifact correction with the different correlation thresholds. Below each artifact-corrected segment the reconstructed horizontal and vertical eye movements are displayed. Using a correlation threshold of 10 % (cf. upper middle panel) the rhythmic brain activity is not adequately modeled. Therefore it projects onto the correlated vertical eye movement topography as can be clearly seen in the 3rd to 5th second of the corresponding VEOG waveform. As a result spatial filtering not only eliminates the artifacts but also causes a severe distortion of the rhythmic brain activity. While brain activity is reduced erroneously in some traces, e.g. P7, Pz or P9, it is added into other channels such as Fp1, Fp2 or F8. With higher correlation thresholds (30 %, 50 %) the distortion gradually decreases (cf. upper right and lower left panel). Using a correlation threshold of 70 % or 90 % the brain activity seems to be modeled adequately. No more distortion is observable. The artifact activity, on the contrary, is reduced to below visibility for every applied correlation threshold. The same EEG segment has already been decomposed by surrogate MSEC in Fig. 3.2. Surrogate MSEC, however, could not completely separate artifact and brain activity.

5.2.2. Example 2: Influence of number of eigenvectors

In the second example the influence of the number of eigenvectors on artifact correction is demonstrated. In the upper left panel of Fig. 5.2 the original EEG segment (4 s, 1–70 Hz, reference FCz) superimposed by a blink and horizontal eye movements in the 2nd and 3rd second is shown. Blink and horizontal eye movement topographies were defined on single prototypes of the respective artifact category. The amplitude threshold was set to 70 µV as the 5-Hz epileptic activity does not exceed a maximum amplitude of about 60 µV. The correlation threshold was established at 50 % giving the best results with a suitable number of eigenvectors. The number of eigenvectors was varied from 1 to 5 corresponding to an explained variance of 77.4 % (1 eigenvector), 85.4 % (2 eigenvectors), 90.1 % (3 eigenvectors), 92.8 % (4 eigenvectors) and 94.9 % (5 eigenvectors). The remaining panels of Fig. 5.2 show the data segment after artifact correction with increasing numbers of eigenvectors. Below each artifact-corrected segment the reconstructed horizontal eye movement and blink waveforms are displayed. The artifact activity is eliminated in each corrected segment. Using 1 eigenvector results in an almost distortion-free artifact correction. With 2 and 3 eigenvectors only minor changes occur. Two differences are marked by arrows in the top middle and top right panel. Using 4 or 5 eigenvectors considerably enhances the noise in the reconstructed HEOG. In the case of 5 eigenvectors noise is also enhanced in the blink waveform. Accordingly, the muscle activity in traces Fp2 and F8 that can already be observed in the original EEG increases. Moreover, noise is introduced into channels F9, F7 and F10. The noise enhancement aggravates with an increasing number of eigenvectors. Correspondingly, the correlation between each artifact topography and the brain signal subspace rises as can be seen in Table 5.1. Thus, the observed noise enhancement seems to be caused mainly by artifact contamination of the estimated brain signal subspace.

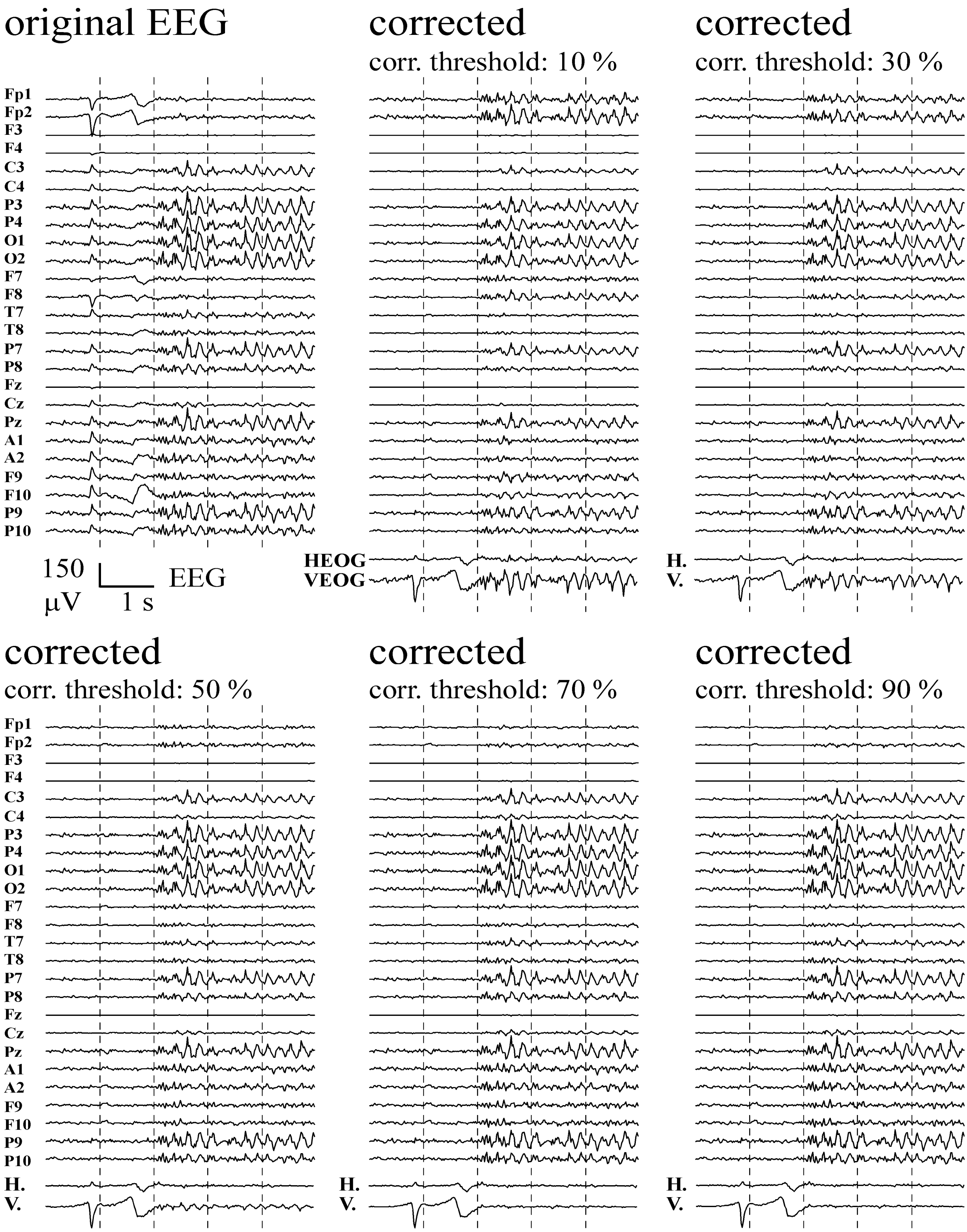

Fig. 5.1: Influence of correlation threshold on artifact correction in the preselection approach. Upper left panel: An EEG segment (5 s, 1-70 Hz) with a blink and an oblique eye movement. Remaining panels: The EEG segment after artifact correction with different correlation thresholds. The reconstructed eye artifacts are depicted in separate traces below each segment.

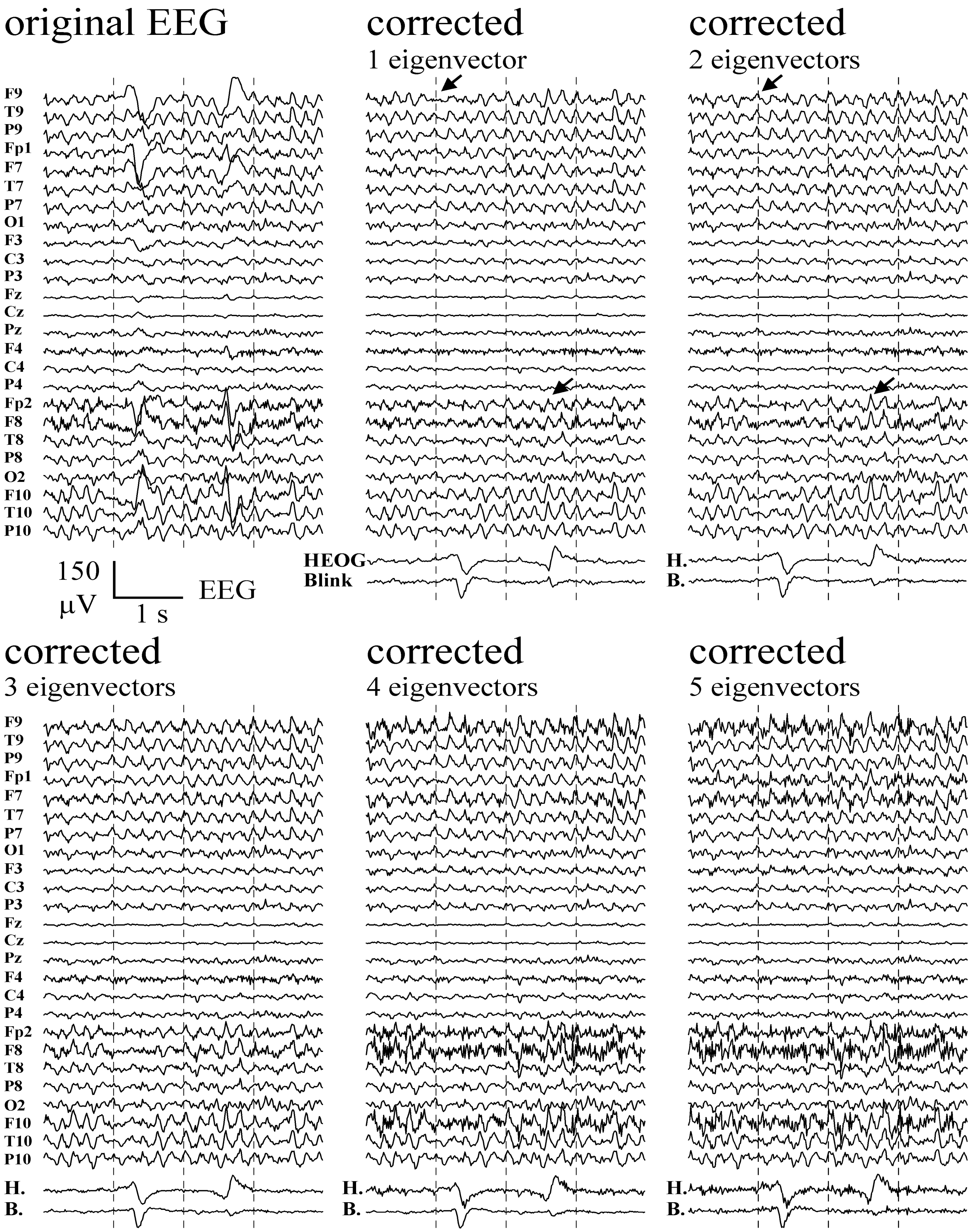

Fig. 5.2: Influence of the number of eigenvectors on artifact correction in the preselection approach. Upper left panel: An EEG segment (4 s, 1-70 Hz) contaminated by a blink and horizontal eye movements. Remaining panels: The EEG segment after artifact correction with different numbers of eigenvectors. The reconstructed eye artifacts are depicted in separate traces below each segment. The arrows mark differences in the corrected data using 1 or 2 eigenvectors.

Table 5.1: Correlation between artifact topographies and the brain signal subspace spanned by different numbers of eigenvectors (cf. Example 2 in chapter 5.2.2 and Fig. 5.2).

	HEOG	Blink
1 eigenvector	3.1 %	14.6 %
2 eigenvectors	70.7 %	14.8 %
3 eigenvectors	85.5 %	72.2 %
4 eigenvectors	99.0 %	80.1 %
5 eigenvectors	99.5 %	98.1 %

5.3. Discussion

The preceding examples illustrate that the preselection approach crucially depends on its parameters. With an adequate choice of amplitude threshold, correlation threshold and number of eigenvectors artifacts are eliminated completely without apparent distortion of brain activity. A reasonable amplitude threshold can be found easily in most cases. Correlation threshold and number of eigenvectors have still to be determined empirically. The parameters, especially the correlation threshold, are hard to estimate automatically. Incorrect thresholds may result in severe distortion of brain activity (Fig. 5.1) or considerable noise enhancement (Fig. 5.2).

Further application of the approach has shown that suitable thresholds generally exist for spontaneous recordings. Theoretical considerations suggest, however, that this may not be the case for raw data containing evoked activity of low SNR. Choosing a low number of eigenvectors in this case event-related signals that correlate with artifact activity may not be adequately modeled and may therefore be distorted. Choosing a higher number of eigenvectors artifact activity of low SNR that remained in the 'artifact-free' subset may contaminate the estimated brain signal subspace and result in noise enhancement. Therefore, spatial filtering with preselection is appropriate to artifact correction when reviewing spontaneous EEG/MEG recordings but seems to be less suitable for event-related continuous data.

Spatial filtering with preselection is very fast. Even for large data matrices, artifact correction requires at most a few seconds on an Intel Pentium III-800 MHz processor using the current C implementation. Spatial filtering with preselection can, therefore, be applied while paging through the segments of an EEG/MEG recording. However, parameters may have to be adjusted manually from one epoch to another if the correlation of artifact and brain activity changes or the number of eigenvectors varies with the state of vigilance. This also limits the application of the method in combination with automated processes such as averaging or mean FFT analysis.

In the next chapter an alternative way of modeling the brain signal subspace is introduced. In contrast to the preselection approach, the novel estimate is based on the whole data segment and is free of parameters that are hard to determine automatically.

6. A new approach: Modeling the signal subspace by SCICA

In this chapter the new concept of SCICA (Ille et al., 2001) to model the brain signal subspace is introduced. The method and an iterative algorithm to perform the SCICA decomposition are described in section 6.1. Some specific algorithmic details are presented separately in the Appendix in section 6.5. In 6.2 spatial filtering with SCICA is applied to simulated data and spontaneous EEG and MEG segments. The examples illustrate the good performance of the approach. In chapter 6.3 the speed of convergence of the presented SCICA algorithm is analyzed. Finally, in section 6.4 the novel approach is discussed.

6.1. Method

SCICA uses the prior knowledge about artifact topographies and combines this with the temporal-statistical strategy of ICA to estimate brain signal topographies from the whole data segment to be corrected. Starting from n a priori determined artifact topographies, the artifact-contaminated data segment is decomposed iteratively into p further components until all waveforms are maximally independent under the spatial constraint. The additionally obtained components represent cerebral activity and unmodeled non-brain activity. In chapter 6.1.1 an iterative algorithm to perform the SCICA decomposition is presented. In chapter 6.1.2 the parameters of the algorithm are discussed in detail.

6.1.1. SCICA algorithm

The basic idea of the SCICA algorithm is summarized in the following steps:

1. *Preprocessing*. Optionally, reduce the dimension of the data matrix $\mathbf{D} \in \Re^{m \times t}$ from $\mathrm{rank}(\mathbf{D})$ to l where $n < l < \mathrm{rank}(\mathbf{D})$. Discard artifact topographies from the matrix of predefined artifact topographies $\mathbf{A} \in \Re^{m \times n}$ (cf. chapter 4.2) that do not contaminate the data $\mathbf{D}_l \in \Re^{m \times t}$ of reduced dimensionality (to ease notation it is assumed here that none of the n artifact topographies is discarded from **A**). Optionally, project the columns of **A** and $\mathbf{D}_l$ into $\mathrm{range}(\mathbf{D}_l)$ to reduce the number of parameters in the following iteration.

Begin iteration

2. Allow for slight perturbation of the n projected artifact topographies to compensate for small deviations from predefined artifact topographies.

3. Find an orthonormal basis of the l-dimensional vector space spanned by the projected data with the first n basis vectors being equal to the perturbed and orthonormalized artifact topographies.
4. Determine $p=l-n$ signal topographies as linear combinations of the orthonormal basis vectors.
5. Determine waveforms corresponding to the current set of artifact and signal topographies.
6. Evaluate independence of waveforms using a suitable ICA contrast function.
7. Repeat steps 2. to 6. with optimized parameters for perturbation and linear combination until the contrast function is optimized, i.e. all waveforms are maximally independent under the spatial constraint.

End iteration

8. After convergence, project the l-n signal topographies back.

The decomposition is performed by direct evaluation of an ICA contrast function. Due to the spatial constraint, elaborate ICA optimization strategies based, for instance, on natural gradient descent or joint diagonalization of cumulant matrices cannot be transferred simply to SCICA. The spatial constraint is realized by keeping the predefined artifact topographies almost unchanged during decomposition except for small perturbations. The brain signal topographies are determined as linear combinations. During optimization the linear combination is sought that yields maximally independent waveforms in combination with the perturbed artifact topographies. During preprocessing, artifact topographies that are irrelevant for the current decomposition are discarded. In order to make the decomposition computationally effective it is also useful to reduce the number of SCICA components to be estimated to the minimum necessary. Below, steps 1-8 of the SCICA algorithm are derived in detail.

Step 1: Preprocessing

Reducing dimension

A useful preprocessing step limiting the number of SCICA components to the minimum necessary is to reduce the dimension of **D**. This can be achieved easily by truncating the SVD expansion of **D** at component l:

$$\mathbf{D}_l = \sum_{i=1}^{l} \mathbf{u}_i \sigma_i \mathbf{v}_i^{\mathrm{T}}$$

with left singular vectors (eigenvectors) $\mathbf{u}_i \in \Re^m$, right singular vectors $\mathbf{v}_i \in \Re^t$, singular values $\sigma_1 \geq \ldots \geq \sigma_r$, $r = \mathrm{rank}(\mathbf{D})$ and $n < l < r$. With a suitable choice of l all relevant activity of $\mathbf{D}$ is still represented in $\mathbf{D}_l$. Range$(\mathbf{D}_l)$ is the l-dimensional subspace in $\Re^m$ spanned by artifact and signal topographies. For the remaining preprocessing steps matrix $\mathbf{C}_\mathrm{o} = (\mathbf{u}_1, \ldots, \mathbf{u}_l) \in \Re^{m \times l}$ is needed whose columns (left singular vectors!) form an orthonormal basis of range$(\mathbf{D}_l)$.

Discarding artifact topographies

Artifact topographies falling below a certain subspace correlation $corr_{\min}$ with $\mathbf{C}_\mathrm{o}$ (e.g. $corr_{\min} = 0.99$) are not considered to contaminate $\mathbf{D}_l$ and are therefore discarded from $\mathbf{A}$. The subspace correlation may be calculated as the cosine of the principal angle between artifact topography and $\mathbf{C}_\mathrm{o}$. Alternatively, it may be derived from the scalar products of the (already normalized) artifact topography $\mathbf{a}_i$ and the l column vectors of $\mathbf{C}_\mathrm{o}$ according to

$$corr_i = \sqrt{\sum_{j=1}^{l} \left(\mathbf{a}_i^\mathrm{T} \mathbf{u}_j\right)^2} \quad .$$

Projection

In order to reduce the parameters in the following iteration, the columns of $\mathbf{A}$ and $\mathbf{D}$ are projected into range$(\mathbf{D}_l)$:

$$\overline{\mathbf{A}} = \mathbf{C}_\mathrm{o}^\mathrm{T} \mathbf{A}$$

$$\overline{\mathbf{D}} = \mathbf{C}_\mathrm{o}^\mathrm{T} \mathbf{D} = \mathbf{C}_\mathrm{o}^\mathrm{T} \mathbf{D}_l = \sum_{i=1}^{l} \sigma_i \mathbf{v}_i^\mathrm{T} \, ,$$

where $\overline{\mathbf{A}} \in \Re^{l \times n}$ is the matrix of projected artifact topographies and $\overline{\mathbf{D}} \in \Re^{l \times t}$ is the projected data matrix. Projecting the columns of $\mathbf{D}$ into range$(\mathbf{D}_l)$ is equivalent to replacing the m originally observed waveforms by their first l basic waveforms. Range$(\overline{\mathbf{D}})$ is the l-dimensional vector space $\Re^l$ spanned by the n projected artifact topographies and the l-n unknown projected signal topographies. In the following iteration we are looking for l-n projected signal topographies in $\Re^l$.

Step 2: Iteration: Perturbation of projected artifact topographies

As the SCICA decomposition may be quite sensitive to small deviations from the predefined artifact topographies, the projected artifact topographies $\overline{\mathbf{A}} = (\overline{\mathbf{a}}_1, \ldots, \overline{\mathbf{a}}_n) \in \Re^{l \times n}$ are slightly perturbed at each iteration. To ensure only small variations, the perturbed artifact topographies in the k^{th} iteration $\overline{\mathbf{A}}^{(k)} = (\overline{\mathbf{a}}_1^{(k)}, \ldots, \overline{\mathbf{a}}_n^{(k)}) \in \Re^{l \times n}$ may not exceed a maximum angle of $\varepsilon_{\max}$ with the projected artifact topographies:

$-\varepsilon_{\max} \leq \measuredangle\left(\overline{\mathbf{a}}_i, \overline{\mathbf{a}}_i^{(k)}\right) \leq \varepsilon_{\max}$ for $i = 1 \ldots n$.

Empirically we have found $\varepsilon_{\max} = 1.5°$ to be a suitable upper boundary. The $\overline{\mathbf{a}}_i^{(k)}$ are derived from the $\overline{\mathbf{a}}_i$ using the current set of perturbation parameters that consists of $n(l\text{-}1)$ angles (cf. Appendix A2 in chapter 6.5.2).

In order to find a particular orthonormal basis of $\Re^l$, i.e. range$(\overline{\mathbf{D}})$, in the next step an orthonormal basis of the perturbed artifact subspace range($\overline{\mathbf{A}}^{(k)}$) is needed. The matrix $\overline{\mathbf{A}}_{\mathrm{o}}^{(k)} \in \Re^{l \times n}$ whose columns form an orthonormal basis of range($\overline{\mathbf{A}}^{(k)}$) may be determined by QR-decomposition of $\overline{\mathbf{A}}^{(k)}$.

Step 3: Iteration: Orthonormal basis of $\Re^l$

In step 3, an orthonormal basis of $\Re^l$ with the first n basis vectors being equal to $\overline{\mathbf{A}}_{\mathrm{o}}^{(k)}$ is sought. This particular basis makes it possible to determine the unknown projected signal topographies as linear combinations while preserving the known projected artifact topographies. The matrix $\overline{\mathbf{C}}_{\mathrm{o}}^{(k)} \in \Re^{l \times l}$ whose columns form an orthonormal basis of $\Re^l$ with $\overline{\mathbf{C}}_{\mathrm{o}}^{(k)}(:, 1:n) = \overline{\mathbf{A}}_{\mathrm{o}}^{(k)}$ may be determined by expanding the basis spanned by the column vectors of $\overline{\mathbf{A}}_{\mathrm{o}}^{(k)}$ as described in Appendix A1 (chapter 6.5.1).

Step 4: Iteration: Projected signal topographies as linear combinations of basis vectors

After $\overline{\mathbf{C}}_{\mathrm{o}}^{(k)}$ has been derived, l-n (projected) signal topographies are determined as linear combinations of the column vectors of $\overline{\mathbf{C}}_{\mathrm{o}}^{(k)}$. Linear combination is accomplished by

$\overline{\mathbf{C}}^{(k)} = \overline{\mathbf{C}}_{\mathrm{o}}^{(k)} \mathbf{R}^{(k)}$.

The column vectors of $\overline{\mathbf{C}}^{(k)} \in \Re^{lxl}$ are the linearly combined topographies. The matrix $\mathbf{R}^{(k)} \in \Re^{l\times l}$ contains the weights of linear combination in the k^{th} iteration. The column vectors $\mathbf{r}_i^{(k)}$ of $\mathbf{R}^{(k)}$ share the following properties:

- $\mathbf{r}_i^{(k)} = \mathbf{e}_i$ for $i = 1 \ldots n$ where $\mathbf{e}_i \in \Re^l$ is the i^{th} canonical vector
- $\varepsilon_{\min} \leq \sphericalangle \left(\operatorname{span}\left\{\mathbf{r}_1^{(k)}, \ldots, \mathbf{r}_{i-1}^{(k)}\right\}, \mathbf{r}_i^{(k)}\right) \leq 180° - \varepsilon_{\min}$ for $i = (n+1) \ldots l$.

Property 1 guarantees that the perturbed and orthonormalized artifact topographies are unchanged by linear combination, i.e. $\overline{\mathbf{C}}^{(k)}(:,1:n) = \overline{\mathbf{C}}_{\mathrm{o}}^{(k)}(:,1:n)$. The last ($l$-$n$) column vectors of $\overline{\mathbf{C}}^{(k)}$ contain the linearly combined (projected) signal topographies in the k^{th} iteration.

Property 2 ensures full column rank of $\mathbf{R}^{(k)}$ and $\overline{\mathbf{C}}^{(k)}$. Whether the column vectors are sufficiently linearly independent in a numerical sense is affected by the particular value of $\varepsilon_{\min}$. In most practical cases ($l \leq 20$) we have found $\varepsilon_{\min} = 10°$ to be a suitable lower boundary using floating point precision. Full column rank of $\overline{\mathbf{C}}^{(k)}$ is required in the next step when the waveforms corresponding to $\overline{\mathbf{C}}^{(k)}$ are whitened.

For construction of the last (l-n) column vectors $\mathbf{r}_i^{(k)}$ of $\mathbf{R}^{(k)}$, refer to Appendix A3 (chapter 6.5.3). The algorithm described there uses the current set of linear combination parameters consisting of 0.5(l-n)(l-n+3) angles to determine the $\mathbf{r}_i^{(k)}$.

Step 5: Iteration: Deriving waveforms

The waveforms corresponding to the set of perturbed and orthonormalized artifact topographies and linearly combined signal topographies $\overline{\mathbf{C}}^{(k)}$ in the k^{th} iteration are derived by

$$\mathbf{S}^{(k)} = \left(\overline{\mathbf{C}}^{(k)}\right)^{-1} \overline{\mathbf{D}},$$

where the rows of $\mathbf{S}^{(k)} \in \Re^{l\times t}$ contain the waveforms. To enable the usage of simpler orthogonal contrast functions in step 6, the waveforms have to be whitened:

$$\mathbf{S}_{\mathrm{W}}^{(k)} = \mathbf{W}^{(k)} \mathbf{S}^{(k)},$$

where $\mathbf{W}^{(k)} \in \Re^{l\times l}$ is the whitening matrix in the k^{th} iteration. The rows of $\mathbf{S}_{\mathrm{W}}^{(k)} \in \Re^{l\times t}$ contain the whitened waveforms. Applying the inverse of $\mathbf{W}^{(k)}$ to $\overline{\mathbf{C}}^{(k)}$

$$\overline{\mathbf{C}}_{\mathrm{W}}^{(k)} = \overline{\mathbf{C}}^{(k)} \left(\mathbf{W}^{(k)}\right)^{-1}$$

yields the topographies $\overline{\mathbf{C}}_{\mathrm{W}}^{(k)} = \left(\overline{\mathbf{c}}_{\mathrm{W},1}^{(k)},\ldots,\overline{\mathbf{c}}_{\mathrm{W},l}^{(k)}\right) \in \Re^{l \times l}$ corresponding to the whitened waveforms as is apparent from $\overline{\mathbf{D}} = \overline{\mathbf{C}}^{(k)}\mathbf{S}^{(k)} = \overline{\mathbf{C}}_{\mathrm{W}}^{(k)}\mathbf{S}_{\mathrm{W}}^{(k)}$. As whitening the waveforms affects the column vectors $\overline{\mathbf{c}}_i^{(k)}$ of $\overline{\mathbf{C}}^{(k)}$, the matrix $\mathbf{W}^{(k)}$ not only has to fulfill the whitening condition

- $E\left\{\mathbf{S}_{\mathrm{W}}^{(k)}\left(\mathbf{S}_{\mathrm{W}}^{(k)}\right)^{\mathrm{T}}\right\} = \mathbf{I}$ but also
- $\mathrm{span}\left(\overline{\mathbf{c}}_{\mathrm{W},1}^{(k)},\ldots,\overline{\mathbf{c}}_{\mathrm{W},n}^{(k)}\right) = \mathrm{span}\left(\overline{\mathbf{c}}_1^{(k)},\ldots,\overline{\mathbf{c}}_n^{(k)}\right)$.

Condition 2 guarantees that the artifact subspace spanned by the first n column vectors of $\overline{\mathbf{C}}^{(k)}$ remains unchanged even if the topographies themselves are modified by whitening. The last (l-n) column vectors of $\overline{\mathbf{C}}_{\mathrm{W}}^{(k)}$ are the final (projected) signal topographies of the k^{th} iteration. For construction of $\mathbf{W}^{(k)}$ see Appendix A4 (chapter 6.5.4).

If $\overline{\mathbf{C}}^{(k)}$ has linearly dependent columns despite step 4, $\mathbf{S}^{(k)}$ can still be calculated by replacing the inverse in the first formula by the pseudo-inverse. In this case, however, the rows of $\mathbf{S}^{(k)}$ are linearly dependent and the whitening matrix $\mathbf{W}^{(k)}$ cannot be determined (see chapter 6.5.4). Consequently, the algorithm has to be restarted with a different initialization, a lower number l of SCICA components, or a higher value of $\varepsilon_{\min}$.

Step 6: Iteration: Evaluation of waveform independence

To assess the independence of the whitened waveforms $\mathbf{S}_{\mathrm{W}}^{(k)} = \left(\mathbf{s}_{\mathrm{W},1}^{(k)},\ldots,\mathbf{s}_{\mathrm{W},l}^{(k)}\right)^{\mathrm{T}}$ in the k^{th} iteration, a contrast function is evaluated. We have investigated the orthogonal fourth-order cumulant-based contrast $\phi_{\mathrm{ICA}}^{\mathrm{o}}$ (Comon, 1994) and the orthogonal log-likelihood contrast $\phi_{\mathrm{L}}^{\mathrm{o}}$ that avoids explicit estimation of the fourth-order statistics. The latter contrast was derived from the non-orthogonal log-likelihood contrast ϕ_{L} of Pham et al. (1992). Comparable to the extended infomax algorithm (Girolami, 1998; Lee et al., 1999) $\phi_{\mathrm{L}}^{\mathrm{o}}$ switches between a sub- and super-Gaussian density function. The switching rule (Lee et al., 1999) is based on the stability analysis of Cardoso and Laheld (1996). Both orthogonal contrast functions, $\phi_{\mathrm{ICA}}^{\mathrm{o}}$ and $\phi_{\mathrm{L}}^{\mathrm{o}}$, are only valid under the whiteness constraint which is fulfilled by whitening the waveforms in step 5. Minimizing either contrast function minimizes the statistical dependence of the waveforms $\mathbf{s}_{\mathrm{W},i}^{(k)}$.

Alternative 1: Fourth-order cumulant-based contrast

The fourth-order auto-cumulant-based contrast function $\phi^{\mathrm{o}}_{\mathrm{ICA}}$ is described in chapter 3.4.4. It is defined as (Comon, 1994):

$$\phi^{\mathrm{o}}_{\mathrm{ICA}} = -\sum_{i=1}^{l} \mathrm{Cum}^2\left[\mathbf{s}^{(k)}_{\mathrm{W,i}}, \mathbf{s}^{(k)}_{\mathrm{W,i}}, \mathbf{s}^{(k)}_{\mathrm{W,i}}, \mathbf{s}^{(k)}_{\mathrm{W,i}}\right] = \min$$

where $\mathrm{Cum}[s,s,s,s] = \mathrm{E}\{s^4\} - 3 + 12\mathrm{E}^2\{s\} - 4\mathrm{E}\{s\}\,\mathrm{E}\{s^3\} + 6\mathrm{E}^4\{s\}$ is the fourth-order auto-cumulant of the non zero-mean random variable s with $\mathrm{E}\{s^2\} = 1$ (whiteness constraint) expressed in terms of expectations. For a definition of cumulants refer to chapter 3.4.7.

Alternative 2: Log-likelihood contrast

From the normalized log-likelihood contrast ϕ_{L} (Pham et al., 1992) described in chapter 3.4.4 the following orthogonal contrast function can be derived (cf. Appendix A5 in chapter 6.5.5):

$$\phi^{\mathrm{o}}_{\mathrm{L}} = -\frac{1}{t}\sum_{j=1}^{t}\sum_{i=1}^{l} \ln \mathrm{p}_i\left(\mathbf{s}^{(k)}_{\mathrm{W},i}(j)\right) = \min$$

where $\mathrm{p}_i\left(\mathbf{s}^{(k)}_{\mathrm{W},i}(j)\right)$ is the pdf of the i^{th} whitened waveform in the k^{th} iteration. For each waveform (here denoted by **s**) either the sub-Gaussian density function

$$\mathrm{p}_{\mathrm{sub}}(\mathbf{s}) = \frac{1}{2}\left(\, \mathrm{N}(1,1) + \mathrm{N}(-1,1)\, \right)$$

or the super-Gaussian density function

$$\mathrm{p}_{\mathrm{super}}(\mathbf{s}) = c \cdot \mathrm{N}(0,1)\,\mathrm{sech}^2(\mathbf{s})$$

is used where $\mathrm{N}(\mu,\sigma^2)$ is the Gaussian density with mean μ and variance σ^2. The normalization constant $c \approx 1.650967$ was determined by numerical integration in Maple V (release 4). The same pdfs are utilized in the extended infomax algorithm (Girolami, 1998; Lee et al., 1999) described in chapter 3.4.4. In Fig. 3.6 they are depicted together with the Gaussian density N(0,1). The switching rule between sub- and super-Gaussian pdf is adopted from the extended infomax algorithm of Lee et al. (1999):

$$\mathrm{sign}\left(\, \mathrm{E}\{\mathrm{sech}^2(\mathbf{s})\}\mathrm{E}\{\, \mathbf{s}^2\} - \mathrm{E}\{\, \tanh(\mathbf{s})\mathbf{s}\,\}\right) = \begin{cases} +1 & \rightarrow \quad \mathrm{p}_{\mathrm{super}}(\mathbf{s}) \\ -1 & \rightarrow \quad \mathrm{p}_{\mathrm{sub}}(\mathbf{s}) \end{cases}.$$

For a derivation of the switching rule refer to chapter 3.4.5.

Step 7: Iteration: Optimization

The angles used to derive the perturbed artifact topographies and the weighting vectors of linear combination are optimized in each iteration applying the simplex algorithm for non-linear optimization by Nelder and Mead (1965) until the chosen contrast function is minimized. In order to minimize a function of x variables, the simplex method compares the function values at $(x+1)$ x-dimensional vertices and replaces the vertex with the highest value by another point. The simplex enclosed by the vertices adapts itself to the local landscape and contracts on to the final (possibly local) minimum. We have employed the simplex implementation of Press et al. (1992). In this implementation the optimization terminates when the contrast function value of the vertex with the minimum value exceeds a particular portion $rtol = (2 - ftol)/(2 + ftol)$ of the contrast function value of the vertex with the maximum value. The fractional convergence tolerance $ftol$ was set to 1.e^{-4} corresponding to $rtol$= 99.99 %.

Step 8: Projecting back the signal topographies

After convergence in the k^{th} iteration, the topography matrix $\overline{\mathbf{C}}_{\mathrm{W}}^{(k)}$ has to be projected back (cf. step 1: projection):

$$\mathbf{C}_{\mathrm{W}}^{(k)} = \mathbf{C}_{\mathrm{o}} \overline{\mathbf{C}}_{\mathrm{W}}^{(k)} \in \Re^{m \times l}.$$

The last p=l-n columns of $\mathbf{C}_{\mathrm{W}}^{(k)}$ contain the topographies of the maximally independent signal components found under the spatial constraint

$$\mathbf{B} = \mathbf{C}_{\mathrm{W}}^{(k)}(:, n+1 : l) \text{ with } \mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_p) \in \Re^{m \times p}.$$

Using the matrix of predefined artifact topographies **A** and the matrix of signal topographies **B**, the spatial filter operator may be composed as outlined in chapter 4.4.

6.1.2. Parameters of the SCICA algorithm

In this chapter the parameters of the SCICA algorithm are presented in detail: the number l of SCICA components, the correlation threshold $corr_{\min}$ used to decide whether a predefined artifact contaminates the current data segment, the maximum allowed perturbation angle of artifact topographies $\varepsilon_{\max}$ and the minimum angle $\varepsilon_{\min}$ ensuring full column rank of the linear combination matrix. The number of components is the only parameter that has to be set for SCICA. It can be estimated automatically. The other parameters may be kept constant. They were set to $corr_{\min} = 0.99$, $\varepsilon_{\max} = 1.5°$ and $\varepsilon_{\min} = 10°$ in the current thesis.

Number of SCICA components

The number l of SCICA components is the main parameter of the SCICA decomposition. It has to be set individually for each decomposed segment. The maximum number of SCICA components is equivalent to the rank of the original data segment. The number of components may be limited by reducing the dimension of the data segment (cf. step 1, chapter 6.1.1), i.e. by truncating the SVD of the original data matrix at component l. If l is set too low, relevant activity in the original data segment may not be considered. If l is chosen too high, the SCICA decomposition may become relatively slow as will be shown in chapter 6.3 without achieving any further apparent improvement of the decomposition. For spontaneous EEG/MEG, the dimension of the data segment may be reduced to the number of left singular vectors (eigenvectors) each explaining at least 1 % of the total data variance. Applying this 1 % rule the number of SCICA components can be determined automatically.

Correlation threshold $corr_{\min}$

Before the SCICA decomposition is calculated, predefined artifact topographies that are not part of the current data segment are discarded (cf. step 1, chapter 6.1.1). An artifact is only considered to contaminate the current epoch if its predefined topography exceeds a minimum subspace correlation $corr_{\min}$ with the data segment of reduced dimension $\mathbf{D}_l$. In the current implementation of SCICA a relatively high threshold $corr_{\min} = 0.99$ is applied. Establishing the threshold too low, brain signals correlating with the artifact above the threshold may suggest that the particular artifact is present in the current epoch. In this case the artifact topography is retained even if the artifact does not occur. Supposing the artifact is not present, the correlated brain signal is partially represented by the artifact topography and is distorted by spatial filtering. On the contrary, setting the correlation threshold too high, an artifact of very low variance in the original data segment $\mathbf{D}$ may not be recognized on condition that it is not completely represented in the epoch of reduced dimension $\mathbf{D}_l$ (depending on l). The corresponding artifact topography is discarded erroneously and the artifact remains uncorrected. The latter problem may be avoided, however, by investigating whether a predefined artifact topography is part of the original data segment $\mathbf{D}$ instead of the epoch of reduced dimension $\mathbf{D}_l$. The correlation threshold is important to exclude predefined topographies automatically from artifact correction without visual control. If only selected epochs are corrected, unnecessary artifact topographies may be discarded manually.

Perturbation angle $\varepsilon_{\max}$

Perturbation of artifact topographies in step 2 of the SCICA decomposition (chapter 6.1.1) seems to be an important issue. Simulations have shown that with a slightly misspecified artifact topography that is kept constant during optimization the originally simulated waveforms may not be completely separable as the simulated signal topographies do not seem to be part of the vector space accessible with the particular artifact topography. Therefore, the artifact topographies are allowed to vary up to a maximum angle $\varepsilon_{\max}$. If $\varepsilon_{\max}$ is chosen too small, not every conceivable variation of artifact topographies may be accounted for. The optimal signal topographies may, however, be sufficiently approximated such that a possibly incomplete separation of waveforms is in the order of magnitude of noise only. If $\varepsilon_{\max}$ is set too high, it may be larger than the (unknown) minimum angle between an artifact and a signal topography. In this case the perturbed artifact topography may represent the correlated signal topography while one of the signal topographies is equivalent to the actual artifact topography. This leads to considerable noise enhancement as the estimated signal subspace that is combined with the original artifact subspace for spatial filtering contains an artifact topography. In the current implementation of SCICA the angle $\varepsilon_{\max}$ is set to a rather small value of 1.5°. Empirically, this threshold has proven to be suitable for spontaneous EEG/MEG recordings.

Linear combination angle $\varepsilon_{\min}$

The angle $\varepsilon_{\min}$ is supposed to ensure full column rank of the linear combination matrix (cf. step 4, chapter 6.1.1). If $\varepsilon_{\min}$ is chosen too small, linear independence may not be achievable in a numerical sense depending on the number l of SCICA components. In the current implementation of SCICA the angle $\varepsilon_{\min}$ was set to 10°. With $\varepsilon_{\min} = 10°$ linear dependence was never observed if up to 15 SCICA components were calculated. For $15 \leq l \leq 20$ linear dependence occurred occasionally, especially for $l = 20$. In most cases it was sufficient, though, to restart the algorithm with a different random initialization. Applying $\varepsilon_{\min} = 10°$, no linear dependence happened in the examples of this thesis. If more than 20 components are calculated, the value of $\varepsilon_{\min}$ should be raised. In our observations, increasing $\varepsilon_{\min}$ even for $l < 20$ does not seem to have any apparent influence on the SCICA decomposition, although the set of possible linear combinations in step 4 of the algorithm is restricted. This is probably due to the whitening operation in step 5 that changes the topographies and the angles between the topographies.

For the fourth-order auto-cumulant contrast, $\varepsilon_{\min}$ may even be omitted if the algorithm is altered slightly. First, property 2 imposed on the column vectors $\mathbf{r}_{n+1}^{(k)},\ldots,\mathbf{r}_l^{(k)}$ of the linear combination matrix $\mathbf{R}^{(k)}$ in step 4 of the algorithm is dropped. Equivalently, the linear combination parameter $\alpha_i^{(k)}$ no longer has to be constrained to [$\varepsilon_{\min}$,180°-$\varepsilon_{\min}$] (cf. Appendix A3 and Table 6.4 in chapter 6.5.3). Instead, the particular vectors $\mathbf{r}_{n+1}^{(k)},\ldots,\mathbf{r}_l^{(k)}$ of the k^{th} iteration are taken into account only if the resulting matrix $\mathbf{R}^{(k)}$ still has full column rank in a numerical sense. Otherwise, they are rejected. Consequently, the number of linearly combined brain signal topographies and, thus, the total number of SCICA components may be reduced in step 4. The waveforms corresponding to the possibly reduced set of topographies are derived in step 5 using the pseudo-inverse instead of the inverse. Whitening is possible in this case because the waveforms, i.e. the rows of $\mathbf{S}^{(k)}$, are linearly independent. Finally, in step 6 the current value of the contrast function $\phi_{\text{ICA}}^{\text{o}}$ is calculated over a possibly reduced number of waveforms. As $\phi_{\text{ICA}}^{\text{o}}$ tends towards $-\infty$, solutions with the full number of waveforms yield (in absolute values) a larger value of $\phi_{\text{ICA}}^{\text{o}}$ and should, thus, be preferred quite naturally during optimization. The outlined change of the algorithm does not seem to be applicable with the log-likelihood contrast $\phi_{\text{L}}^{\text{o}}$ that tends towards zero and is, therefore, expected to converge into a useless solution without any brain signal topography.

6.2. Examples

Spatial filtering with SCICA was implemented in C (Microsoft Visual C++ 6.0) and was applied to simulated data as is demonstrated in chapter 6.2.1. The simulation shows the good performance of the decomposition and confirms that it is not influenced by the spatial correlation between artifact and signal subspace. SCICA was also successfully tested on spontaneous EEG and MEG data as is illustrated in chapters 6.2.2 and 6.2.3.

For the simulated data, the quality of artifact correction was quantified. For the real EEG and MEG segments, artifact correction was again evaluated by thorough visual comparison of the epoch before and after artifact correction applying the criteria defined in the first paragraph of chapter 5.2. To ease visual comparison overplots of the original and corrected segments are shown.

Before artifact correction the data dimension was always reduced (cf. step 1 of chapter 6.1.1). For real data, only eigenvectors explaining at least 1 % of the total data variance were retained (1 % rule). For simulated data, the number of eigenvectors was set to the rank of the simulated epoch (rank rule).

6.2.1. Applying SCICA to simulated data

This example uses simulated EEG data to show that the SCICA decomposition is capable of recovering the original independent waveforms from a simulated mixture of waveforms. It also illustrates that SCICA does not depend on the spatial correlation between artifact and signal subspace. The completeness of artifact removal and the distortion of signal activity are quantified. The small quantified distortion is shown to be in the noise range of continuous EEG recordings.

We have simulated three 6-s data sets with a low (10 %), medium (50 %) and high (90 %) spatial correlation between artifact and signal subspace using dipoles in a 4-shell spherical head model. In Fig. 6.1 the simulation is depicted. Dipoles 1 and 2 and their corresponding waveform represent a blink artifact in the 2nd and 5th second. The artifact dipoles are equal in each simulation. Dipoles 3 and 4 model the brain signal subspace. They have been determined randomly. The subspace (plane) they are spanning has the desired correlation of 10 %, 50 % or 90 % with the artifact subspace. The 6-s sawtooth and sine waveforms of dipoles 3 and 4 are unrealistic but have been chosen as they are largely independent.

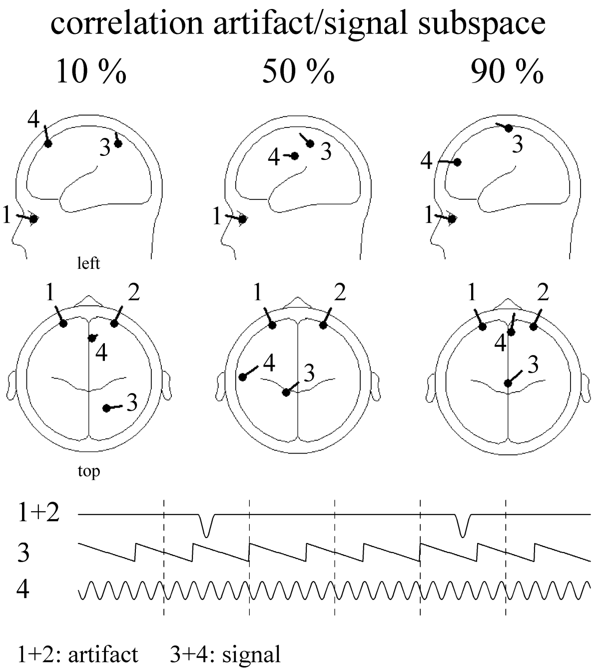

Fig. 6.1: Dipole simulation of three 6-s data sets with a low (10 %), medium (50 %) and high (90 %) correlation between artifact and signal subspace. Dipoles 1 and 2 model the artifact subspace. Dipoles 3 and 4 represent the brain signal subspace. The waveforms are largely independent.

In row (A) of Fig. 6.2 the three simulated data sets are displayed. The electrode configuration (Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, Fz, Cz, Pz, A1, A2, F9, F10, P9, P10) includes the standard 10-20 electrodes plus 3 infero-temporal electrodes on either side of the head. No noise has been added to the simulated data.

In row (B) of the same figure overplots of the original segments (red) and the epochs after spatial filtering with SCICA (black) applying the fourth-order cumulant contrast are shown. The blinks in the 2nd and 5th second are removed in each data set without apparent distortion of the signal activity. The artifact topography was defined in a separate epoch containing a simulated blink only. Before artifact correction the data dimension was reduced to 3 eigenvectors (rank rule).

Finally, in row (C) the three equally scaled SCICA waveforms are depicted. Waveform 1 corresponds to the predefined blink topography, waveforms 2 and 3 represent the signal activity estimated by SCICA. Irrespective of the spatial correlation between artifact and signal subspace the original waveforms have been recovered by SCICA up to a tiny residual signal waveform superimposed on each blink waveform. This residual signal is not visible in the displayed blink waveforms (SCICA waveforms 1) as its average peak-to-peak amplitude amounts to less than 0.41 % of the maximum blink amplitude only (cf. Table 6.1). To account for initialization effects, average values over 100 randomly initialized repetitions of the SCICA decompositions are given. The peak-to-peak signal amplitudes in SCICA waveform 1 were calculated as the difference between maximum and minimum amplitude in the 3rd and 4th second. The maximum blink amplitude in SCICA waveform 1 was determined as the maximum amplitude of the absolute of the whole time series.

Of course, the residual signal is also part of the estimated artifact activity that is subtracted from the original data to obtain the corrected data. The peak-to-peak amplitudes of the residual signal in the subtracted data were determined in the 3rd and 4th second in channel Fp1 where apart from channel Fp2 the weights of the predefined blink topography are maximum, i.e. SCICA waveform 1 is maximally enhanced. On average, the peak-to-peak amplitudes of the residual signal in the subtracted data are less than 0.36 μV (N=100) (cf. Table 6.1). Amplitude changes in this order of magnitude are too small to be visible in panels (B) of Fig. 6.2. Moreover, they lie in the recording noise range of continuous EEG data which may be up to 2 μV peak-to-peak (Nuwer et al., 1999). The quantified peak-to-peak amplitudes can thus be regarded as irrelevant. Dividing the absolute mean peak-to-peak amplitude by the corresponding relative mean peak-to-peak amplitude using the values given in Table 6.1 (with higher precision), the maximum blink amplitude in the subtracted data can be estimated to 86.9 μV in each case. This value is equivalent to the maximum simulated blink amplitude.

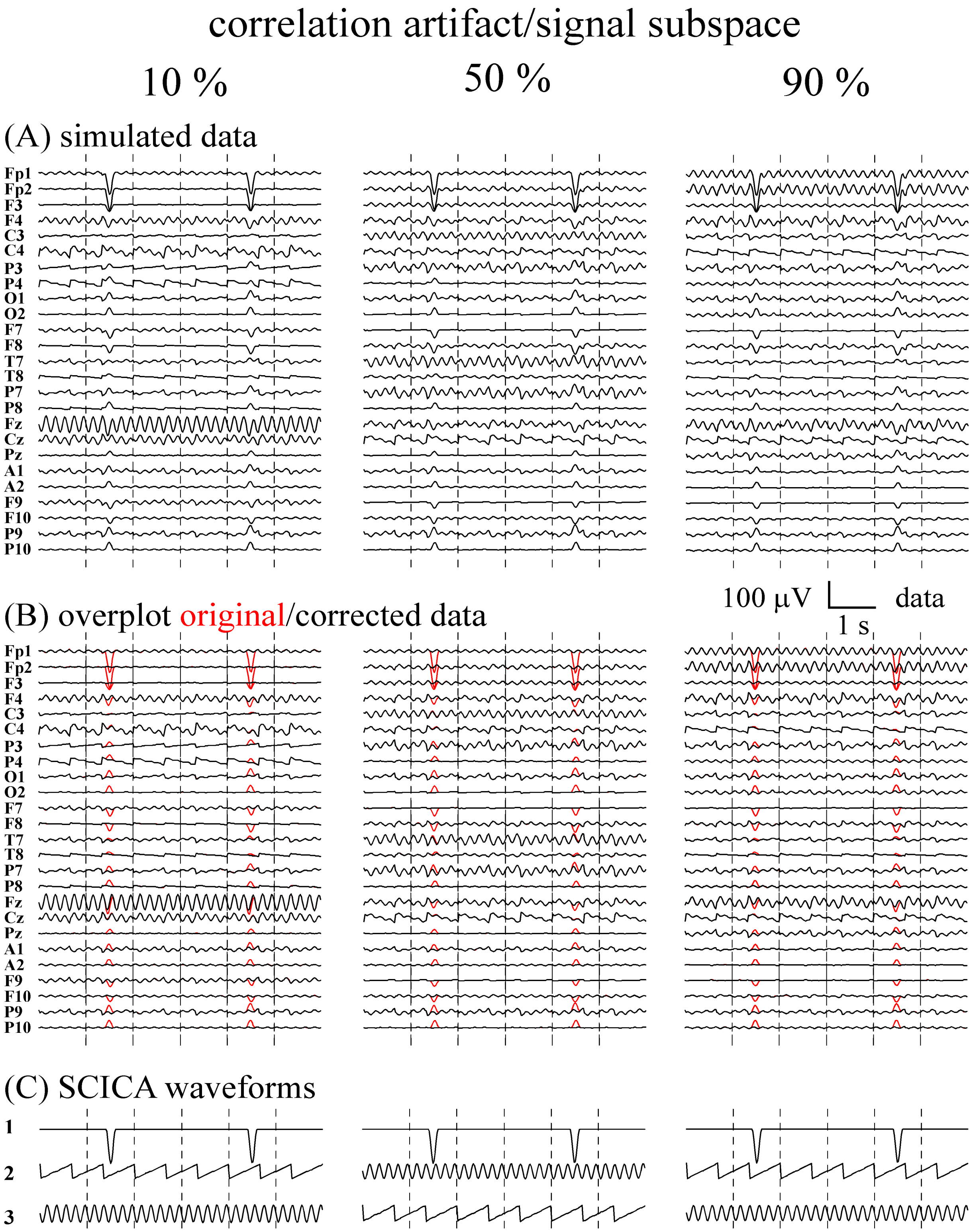

Fig. 6.2: Applying spatial filtering with SCICA to simulated data. (A) The three simulated 6-s data sets with a low (10 %), medium (50 %) and high (90 %) correlation between artifact and signal subspace derived from the dipole simulation in Fig. 6.1. (B) The same segments after spatial filtering. (C) The waveforms recovered by SCICA. Waveform 1 corresponds to the predefined artifact topography.

Table 6.1: Peak-to-peak signal amplitude in SCICA waveform 1 (cf. Fig. 6.2) in percent of maximum blink amplitude in SCICA waveform 1 and peak-to-peak amplitude of the subtracted signal for a correlation between artifact and signal subspace of 10 %, 50 % and 90 %. Mean values and standard deviations determined over N=100 randomly initialized repetitions of the particular decomposition are shown.

	peak-to-peak signal amp. in percent of max. blink amp. in SCICA waveform 1	peak-to-peak amp. of subtracted signal
10 %	0.41 % ± 0.20 %	0.36 μV ± 0.17 μV
50 %	0.40 % ± 0.19 %	0.34 μV ± 0.16 μV
90 %	0.31 % ± 0.13 %	0.27 μV ± 0.11 μV

In order to quantify the completeness of artifact removal, the rank of the subspace spanned by the SCICA corrected data is determined. If the blinks were not completely removed, a rank of three would be expected. The rank was found to be two in each case (cf. Table 6.2). Thus, the blinks are completely eliminated confirming the visual impression in Fig. 6.2 (B). The rank was determined as the number of non-zero singular values. Singular values falling below the tolerance $\max(m,t)\cdot\max(\text{singular value})\cdot\text{eps}$ were set to zero (Dongarra et al., 1979) with number of channels m, number of samples t and the floating point relative accuracy eps=1.192092896e^{-07}.

To quantify the amount of topographic distortion introduced by a residual signal in the blink waveform of the above-mentioned order of magnitude, the angle between the simulated signal subspace spanned by dipole topographies 3 and 4 and the rank 2 subspace spanned by the SCICA corrected data is calculated. The angle is determined as the largest principal angle between the two subspaces (Golub and van Loan, 1996). As can be seen in Table 6.2 the average subspace angles are below 0.48° (N=100) corresponding to a subspace correlation (cosine of subspace angle) of more than 99.99 %.

Table 6.2: Rank of SCICA corrected data and angle between simulated signal subspace and subspace spanned by SCICA corrected data for a correlation between artifact and signal subspace of 10 %, 50 % and 90 %. Mean values and standard deviations determined over N=100 randomly initialized repetitions of the decomposition are shown.

	rank of SCICA corrected data	angle between simulated signal subspace and subspace spanned by SCICA corrected data
10 %	2 ± 0	0.41° ± 0.22°
50 %	2 ± 0	0.48° ± 0.25°
90 %	2 ± 0	0.17° ± 0.07°

The three simulated data sets were also decomposed by ICA. ICA analysis was performed with both Matlab implementations of the JADE and the extended infomax algorithm (see chapter 3.4.5). Both ICA decompositions resulted in a comparable residual signal in the blink waveform.

6.2.2. Applying SCICA to EEG data

In the following three examples spatial filtering with SCICA is applied to spontaneous EEG segments. The presented data contain relevant brain signal activity contaminated by eye movements, blinks or cardiac artifacts. The examples demonstrate the good performance of the approach on real EEG data. They also show that the fourth-order cumulant-based contrast and the log-likelihood contrast presented in chapter 6.1.1 (step 6) perform equally well.

Example 1: Removing blink and eye movement from EEG by SCICA

In the upper left panel of Fig. 6.3 an EEG epoch (10 s, 1-70 Hz, common reference FCz) recorded during an epileptic seizure is shown. The 5-Hz epileptic activity is superimposed by a blink and a horizontal eye movement in the 4th and 5th second. The same segment after artifact correction with SCICA using the log-likelihood contrast function is depicted in the upper right panel. The artifacts are completely removed without apparent distortion of the periodic activity. The SCICA waveforms corresponding to horizontal eye movement and blink are displayed below the corrected EEG. Before artifact correction the data dimension was reduced to 6 eigenvectors (1 % rule). The topography of horizontal eye movement and blink were defined on single prototypes of the respective artifact category. The overplot of original EEG in red and corrected EEG in black in the lower left panel illustrates the good correspondence of the pathological brain activity before and after artifact correction. In the lower right panel the difference waveforms between original and corrected EEG, i.e. the estimated artifact signals at each electrode, are shown. In Fig. 5.2 a portion of the same EEG segment (3rd to 6th second) has already been corrected by spatial filtering with preselection. Choosing an adequate number of eigenvectors for preselection, both approaches perform equally well.

Example 2: Removing eye movement and cardiac artifacts from EEG by SCICA

In the left upper panel of Fig. 6.4 an EEG segment (6 s, 1-70 Hz, referenced to the average of F3 and F4) recorded with eyes closed is shown. The epoch contains normal alpha rhythm and is contaminated by cardiac artifacts and on-going small vertical movements of the eyes under the closed eye-lids. The upper right panel depicts the same segment after spatial filtering with

SCICA using the fourth-order cumulant-based contrast. The reconstructed vertical eye movements and cardiac activity are displayed below the corrected EEG. The vertical eye movement topography was derived from a single prototype. The EKG topography was defined on the averaged QRS-complex of several hundreds of cardiac cycles. Before artifact correction the data dimension was reduced to 8 eigenvectors (1 % rule). Spatial filtering removes the artifacts to below visibility without obvious distortion of the alpha rhythm. This is also illustrated in the overplot of original (red) and corrected EEG (black) in the lower left panel. The estimated artifact signals at each electrode are shown in the lower right panel.

Example 3: Removing blink and eye movement from EEG by SCICA

The upper left panel of Fig. 6.5 depicts another EEG segment (6 s, 1-70 Hz, common reference FCz) contaminated by blinks and a horizontal eye movement. The artifacts coincide temporally with pathological delta bursts. In the upper right panel the same segment is shown after artifact correction by SCICA applying again the log-likelihood contrast function. The eye artifacts are suppressed without recognizable distortion of the pathological delta activity. The reconstructed horizontal eye movement and blink waveforms are displayed below the corrected EEG. The artifact topographies were defined on single prototypes of the respective artifact category. Before artifact correction the data dimension was reduced to 7 eigenvectors (1 % rule). The overplot of the original EEG in red and the corrected data in black shown in the lower left panel reveals that not only the delta bursts but also the remaining brain signal activity is almost completely maintained by spatial filtering with SCICA. Finally, in the lower right panel the estimated artifact signal at each electrode is depicted.

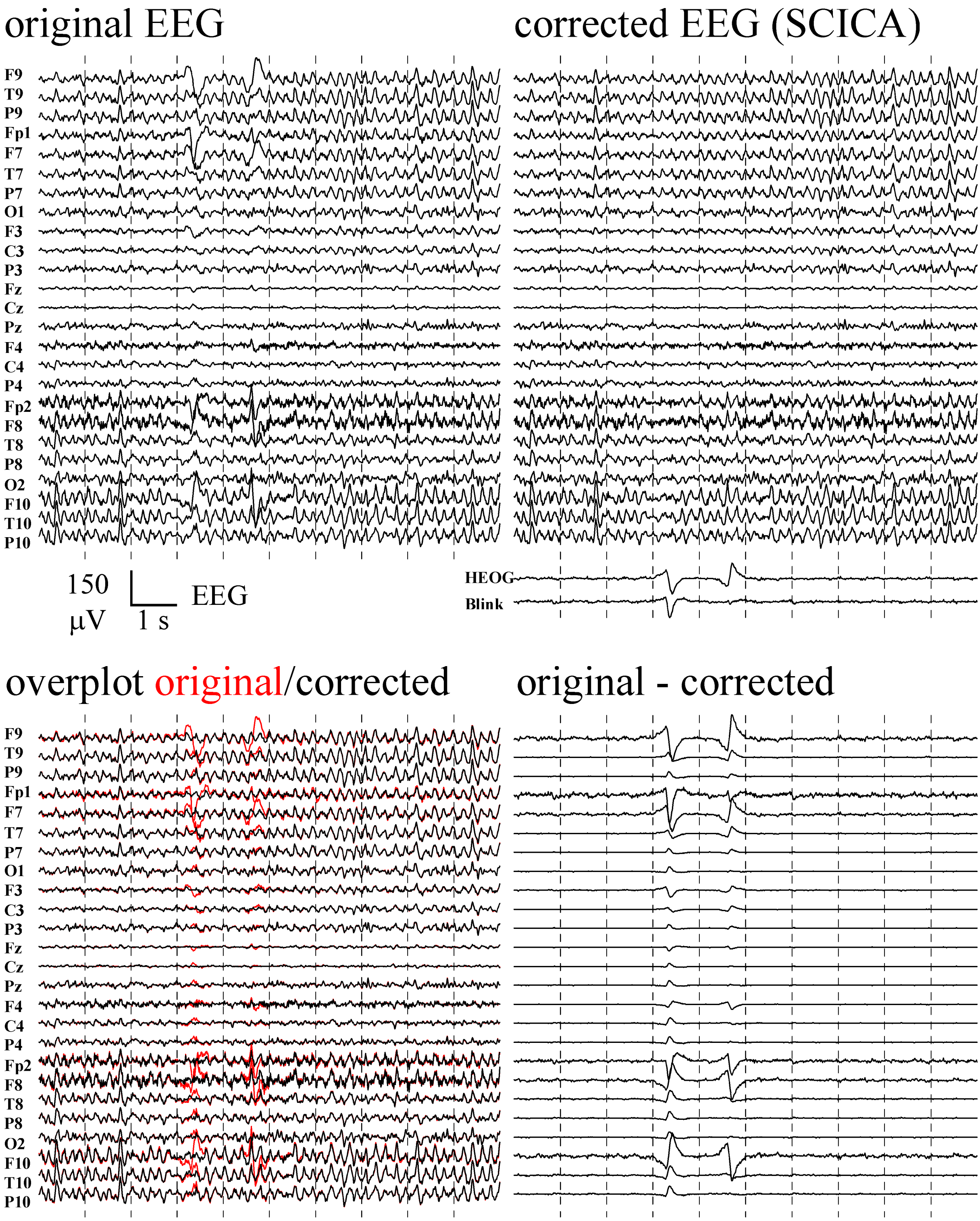

Fig. 6.3: Applying spatial filtering with SCICA to spontaneous EEG data (cf. Example 1 in chapter 6.2.2). Upper left panel: The original EEG segment (10 s, 1-70 Hz) contaminated by a horizontal eye movement and a blink. Upper right panel: The same epoch after artifact correction using the log-likelihood contrast. Below the corrected data the reconstructed horizontal eye movement and blink waveforms are shown. Lower left panel: Overplot of the original (red) and the corrected (black) EEG segment. Lower right panel: Difference between original and corrected EEG segment.

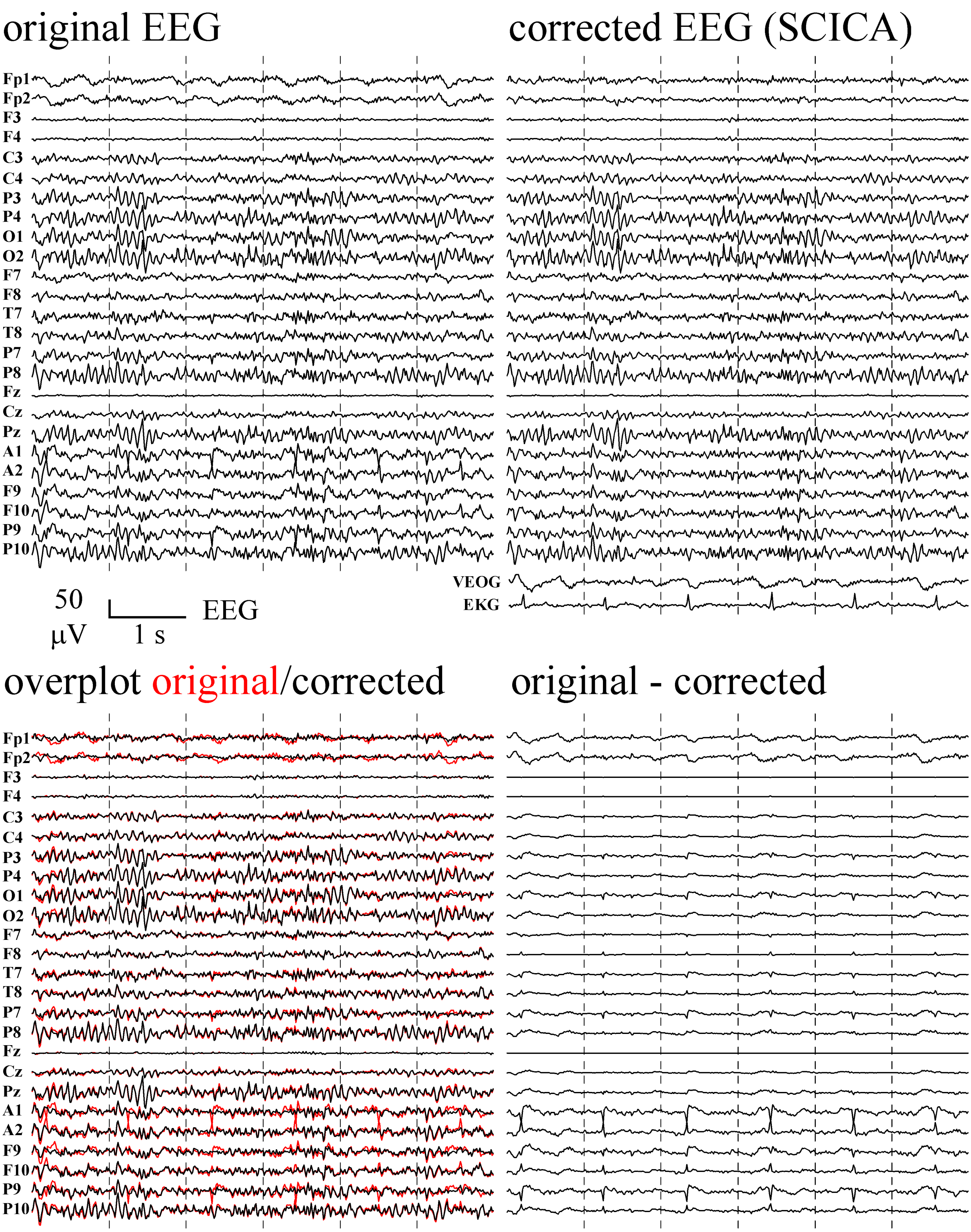

Fig. 6.4: Applying spatial filtering with SCICA to spontaneous EEG data (cf. Example 2 in chapter 6.2.2). Upper left panel: The original EEG segment (6 s, 1-70 Hz) contaminated by vertical eye movements and cardiac artifacts. Upper right panel: The same epoch after artifact correction using the fourth-order cumulant-based contrast. Below the corrected data the reconstructed vertical eye movements and EKG are shown. Lower left panel: Overplot of the original (red) and the corrected (black) EEG segment. Lower right panel: Difference between original and corrected EEG segment.

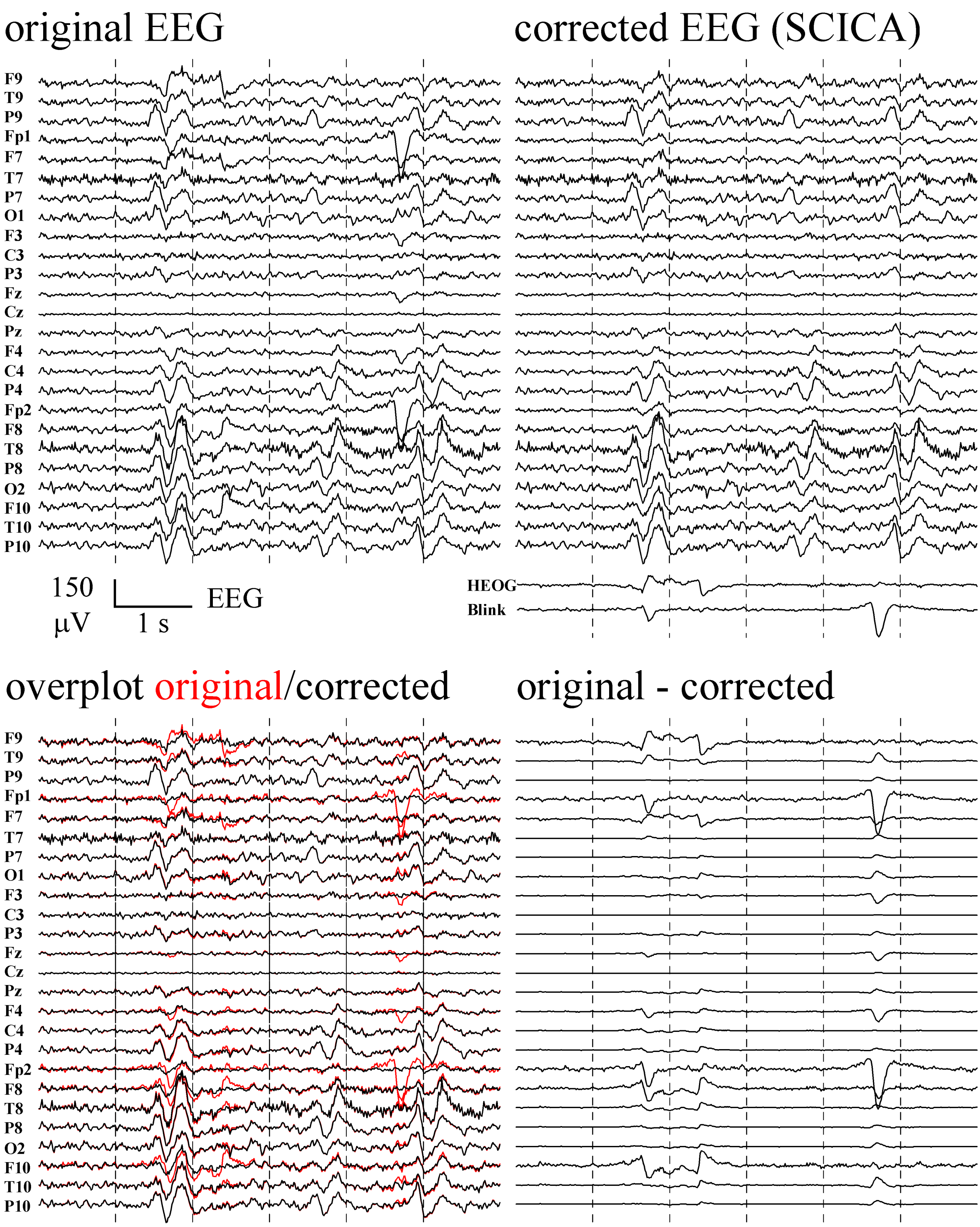

Fig. 6.5: Applying spatial filtering with SCICA to spontaneous EEG data (cf. Example 3 in chapter 6.2.2). Upper left panel: The original EEG segment (6 s, 1-70 Hz) contaminated by blinks and a horizontal eye movement. Upper right panel: The same epoch after artifact correction using the log-likelihood contrast. Below the corrected data the reconstructed horizontal eye movement and blink waveforms are shown. Lower left panel: Overplot of the original (red) and the corrected (black) EEG segment. Lower right panel: Difference between original and corrected EEG segment.

6.2.3. Applying SCICA to MEG data

In the subsequent three examples spatial filtering with SCICA is applied to spontaneous MEG segments. The presented data again contain relevant brain signal activity contaminated by eye movements, blinks or cardiac artifacts. The examples demonstrate the good performance of the approach on real MEG data. They further confirm that the fourth-order cumulant-based contrast and the log-likelihood contrast (step 6, chapter 6.1.1) perform equally well. The MEGs were recorded with a 122-channel whole-head planar gradiometer system (Neuromag Ltd.) in a magnetically shielded room. For clarity, in each example only a subset of relevant sensors measuring the tangential magnetic field gradient along top-bottom meridians are considered. The sensors are named by equivalent EEG electrode labels.

Example 1: Removing cardiac artifacts from MEG by SCICA

The upper left panel of Fig. 6.6 shows an MEG segment (6 s, 1-70 Hz) that is heavily contaminated by cardiac activity. A subset of 32 left-hemispheric and midline sensors is shown in a longitudinal arrangement. The upper right panel depicts the same MEG segment after artifact correction using SCICA with the fourth-order cumulant-based contrast function. The cardiac potentials are corrected without apparent distortion of the theta activity and the delta bursts in the 3rd and 5th second that are most prominent over left centro-parietal sensors. Below the corrected MEG, the SCICA waveform corresponding to cardiac activity is displayed. Before artifact correction the dimensionality of the data was reduced to 9 eigenvectors (1 % rule). The topography of the cardiac artifact was defined on the averaged QRS-complex of several hundreds of cardiac cycles. The overplot of original (red) and corrected (black) MEG in the lower left panel demonstrates the good correspondence of brain activity before and after artifact correction. In the lower right panel the difference waveforms between original and corrected MEG, i.e. the estimated artifact signals at each sensor, are shown.

Example 2: Removing blinks and eye movements from MEG by SCICA

An MEG segment (10 s, 1-70 Hz) containing a series of voluntary horizontal eye movements and blinks is depicted in the upper left panel of Fig. 6.7. Only a subset of 34 frontal to central sensors is displayed in a transversal arrangement. The same MEG epoch after spatial filtering with SCICA using the log-likelihood contrast is shown in the upper right panel. The reconstructed horizontal eye movements and blinks are depicted in separate traces below the corrected MEG. The artifact topographies were defined on single prototypes of the respective artifact category. Before spatial filtering the dimensionality of the data was reduced to 13 eigenvectors (1 % rule). The pronounced brain signal activity in the theta and alpha range over left temporal sensors (FC9, T9) remains unaffected by spatial filtering as is also illustrated in the overplot of original (red) and corrected (black) MEG in the lower left panel. The lower right panel depicts the estimated artifact signals at each sensor.

Example 3: Removing cardiac artifacts from MEG by SCICA

The upper left panel of Fig. 6.8 shows another 6-s MEG epoch contaminated by cardiac artifacts. Only a subset of 33 left-hemispheric and central sensors is displayed in a longitudinal arrangement. The MEG was filtered between 0.1 and 70 Hz to reveal the slow delta activity over deep left frontal, temporal and parietal sensors caused by an ischemic stroke in the territory of the left medial cerebral artery. The same MEG epoch after artifact correction with SCICA again using the fourth-order cumulant-based contrast is shown in the upper right panel. After elimination of the cardiac artifacts the pathological slow activity is much more apparent. Below the corrected data, the reconstructed EKG waveform is displayed. The EKG topography was defined on the averaged QRS-complex of several hundreds of cardiac cycles. Before artifact correction the dimensionality of the data segment was reduced to 11 eigenvectors (1 % rule). In the lower left panel the overplot of original MEG in red and corrected MEG in black is shown. It illustrates that the brain signal activity is almost undistorted by artifact correction. In the lower right panel the estimated cardiac artifacts at each sensor are displayed. Comparison with the overplot reveals that the non-cardiac activity between the 3rd and 4th cardiac cycle in channels FC9, T9 or CP9, for instance, does not introduce any relevant distortion.

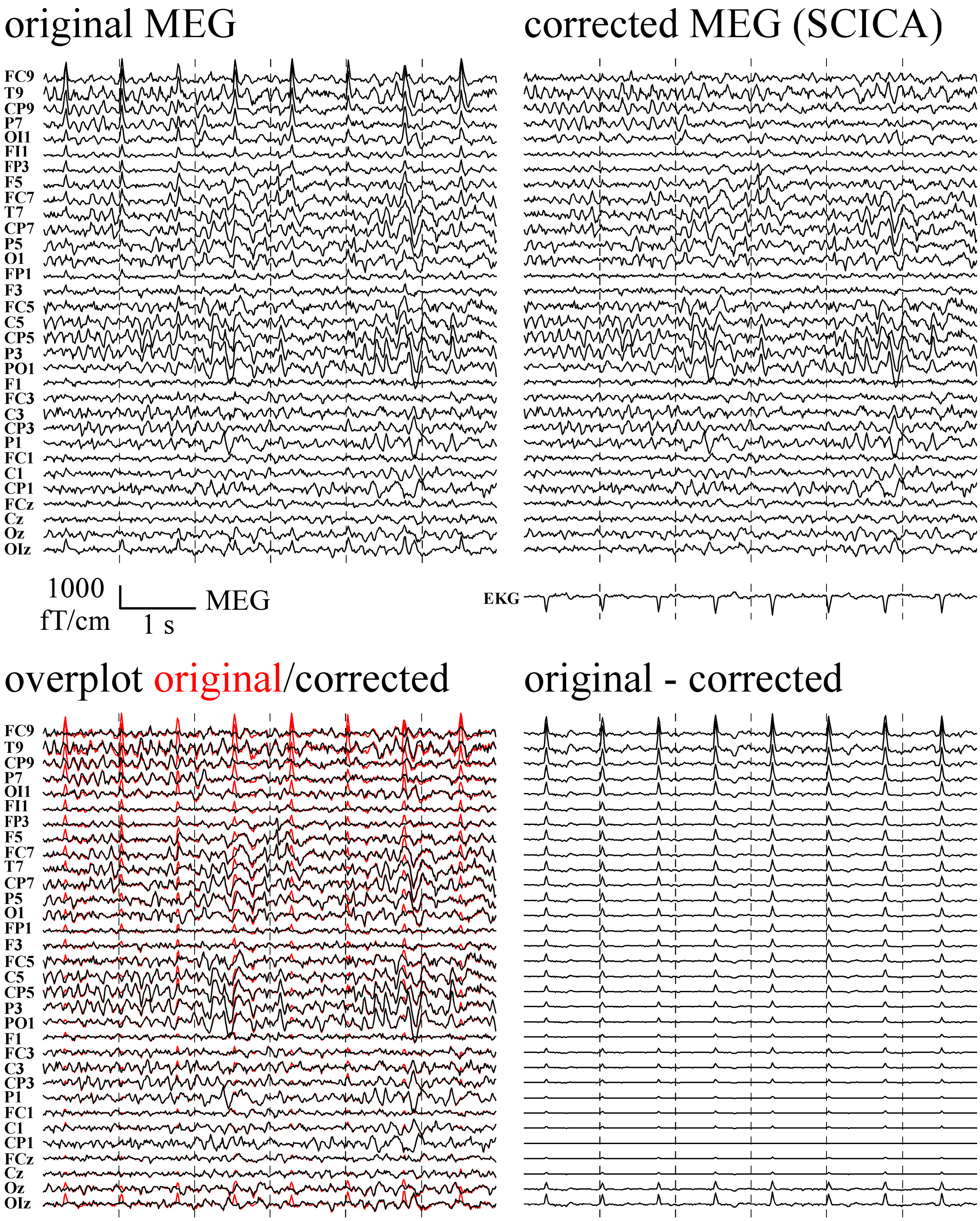

Fig. 6.6: Applying spatial filtering with SCICA to spontaneous MEG data (cf. Example 1 in chapter 6.2.3). Upper left panel: The original MEG segment (6 s, 1-70 Hz) that is heavily contaminated by cardiac artifacts. Upper right panel: The same epoch after artifact correction using the fourth-order cumulant-based contrast. Below the corrected data the reconstructed EKG is shown. Lower left panel: Overplot of the original (red) and the corrected (black) MEG segment. Lower right panel: Difference between original and corrected MEG segment.

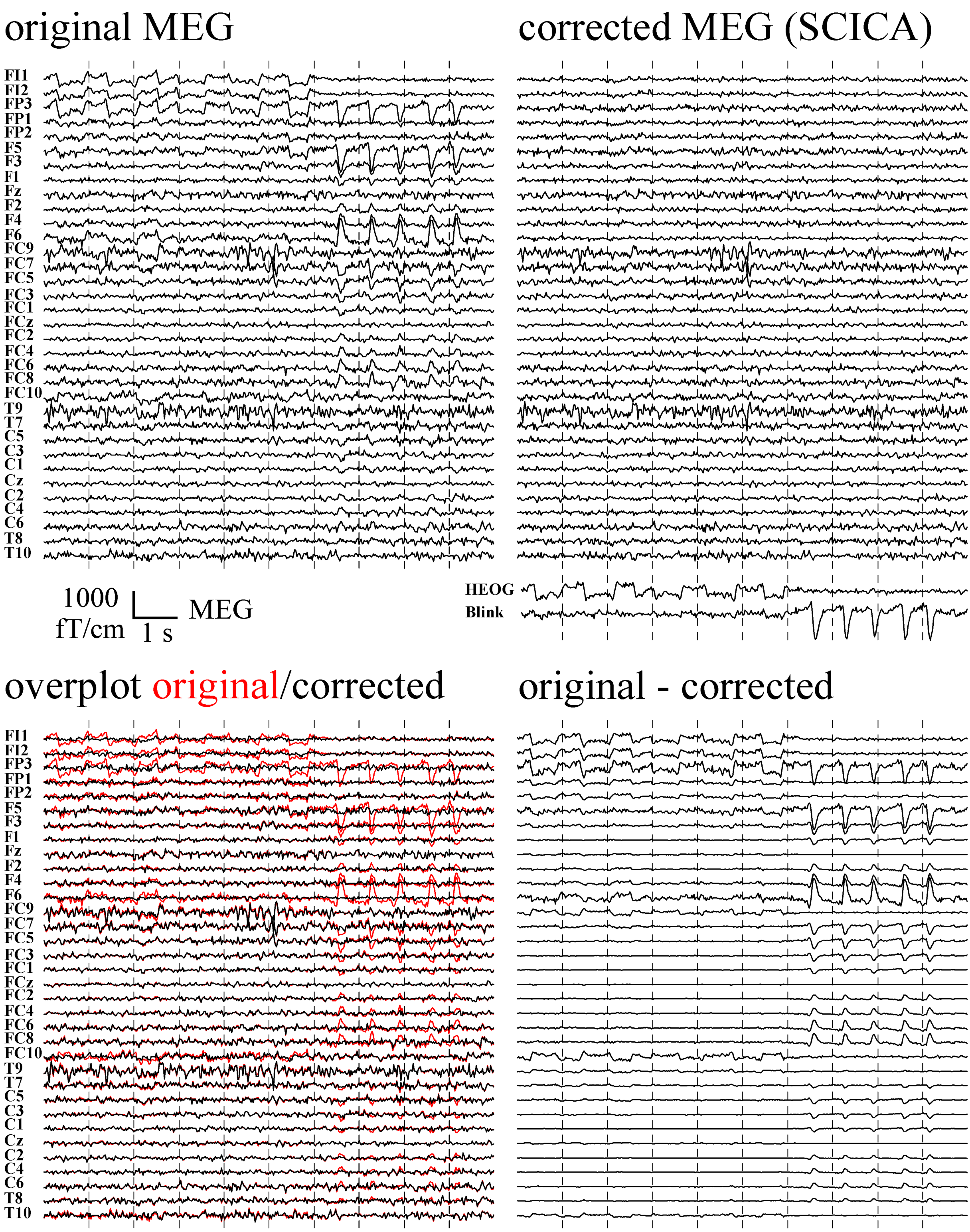

Fig. 6.7: Applying spatial filtering with SCICA to spontaneous MEG data (cf. Example 2 in chapter 6.2.3). Upper left panel: The original MEG segment (10 s, 1-70 Hz) contaminated by horizontal eye movements and blinks. Upper right panel: The same epoch after artifact correction using the log-likelihood contrast. Below the corrected data the reconstructed horizontal eye movements and blinks are shown. Lower left panel: Overplot of the original (red) and the corrected (black) MEG segment. Lower right panel: Difference between original and corrected MEG segment.

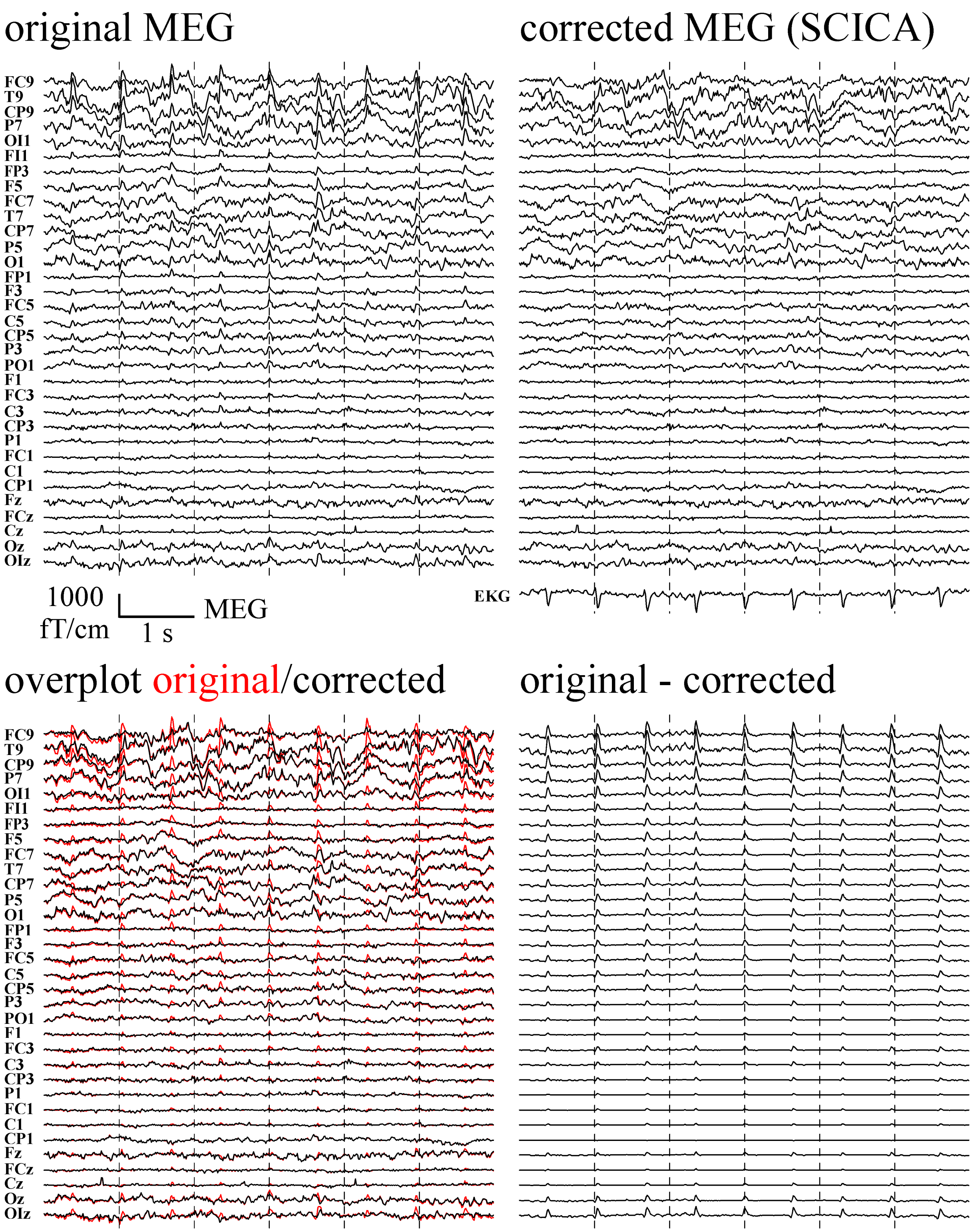

Fig. 6.8: Applying spatial filtering with SCICA to spontaneous MEG data (cf. Example 3 in chapter 6.2.3). Upper left panel: The original MEG segment (6 s, 0.1-70 Hz) contaminated by cardiac artifacts. Upper right panel: The same epoch after artifact correction using the fourth-order cumulant-based contrast. Below the corrected data the reconstructed EKG is shown. Lower left panel: Overplot of the original (red) and the corrected (black) MEG segment. Lower right panel: Difference between original and corrected MEG segment.

6.3. Speed of convergence

As the SCICA decomposition is an iterative process it is not as fast as the preselection approach. A decomposition in the current implementation in C requires a few seconds to several minutes on an Intel Pentium III-800 MHz processor until convergence is achieved. In this chapter the factors influencing the duration of a single iteration and the total number of iterations are analyzed. It is shown that the fourth-order auto-cumulant contrast is considerably faster than the log-likelihood contrast and concluded that a total of 15 to 20 SCICA components should not be exceeded to allow for a computationally efficient decomposition.

In Fig. 6.9 the average duration of a single iteration, i.e. steps 2 to 6 of the algorithm described in chapter 6.1.1, is displayed for different numbers of SCICA components and time samples applying either the fourth-order auto-cumulant contrast or the log-likelihood contrast. The time measurement was performed for six different EEG segments that were increased to comprise 2000, 3000 and 4000 samples corresponding to an epoch of 10 s, 15 s and 20 s with a sample rate of 200 Hz. The number of SCICA components (l in the above algorithm) was set to 5, 10, 15 and 20 including 2 artifact topographies in each decomposition. The elapsed times were obtained as the averages over 100 iterations using the system's high resolution performance counter to measure the duration of a single iteration. The figures show that the mean time required for one iteration increases with a growing number of samples and SCICA components. For a high number of samples, enhancing the number of SCICA components results in a stronger rise of the time per iteration than for a low number of samples.

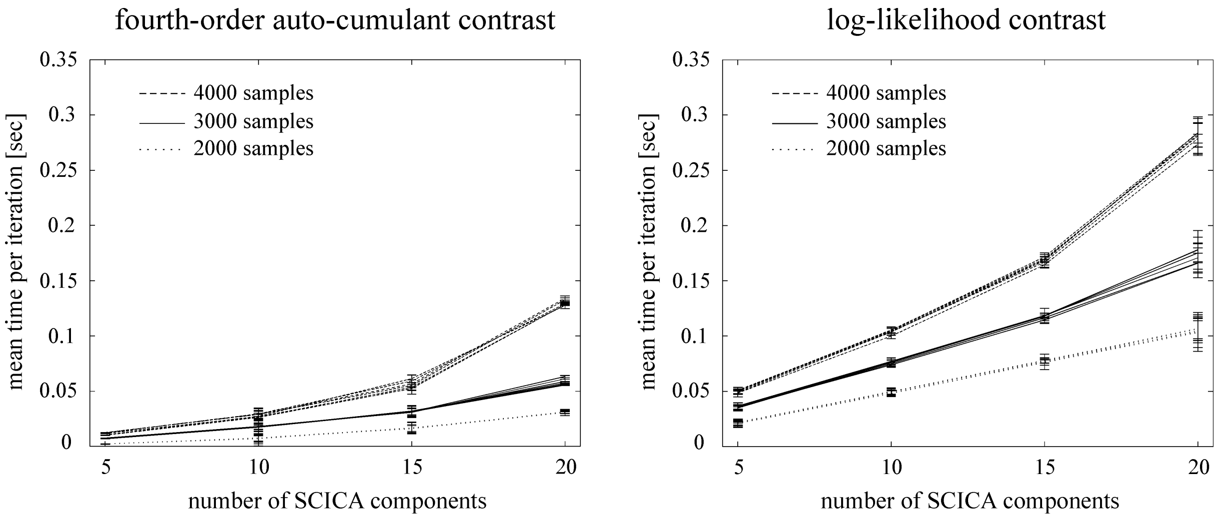

Fig. 6.9: Mean time in seconds for one pass through the SCICA algorithm depending on the number of components and time samples applying either the fourth-order auto-cumulant or the log-likelihood contrast. Mean values and standard deviations of 100 subsequent iterations are displayed.

The fourth-order auto-cumulant contrast is considerably faster than the log-likelihood contrast. Separate time measurement for steps 2 to 6 of the SCICA algorithm reveals that the cumulant contrast spends 70 % (± 12 %) of the time per iteration in step 5 with the matrix multiplications to obtain $\mathbf{S}^{(k)}$ and $\mathbf{S}_{\mathrm{W}}^{(k)}$ and 23 % (± 9 %) in step 6. The likelihood contrast, on the contrary, uses 81 % (± 8 %) of the time per iteration in step 6 with calculation of the switching condition and the contrast function itself and further 18 % (± 9 %) in step 5. The indicated values were determined as the ratios of the mean time per single step (N=100 iterations) and the mean time per iteration (N=100 iterations) averaged over the 72 observations of Fig. 6.9 (4 different numbers of SCICA components times 3 different numbers of samples times 6 different EEG segments).

With an increasing number of SCICA components not only the duration of one iteration, but also the total number of iterations rises as the amount of parameters that has to be estimated by the Nelder and Mead simplex algorithm grows. The relationship between number of SCICA components and parameters is depicted in Fig. 6.10.

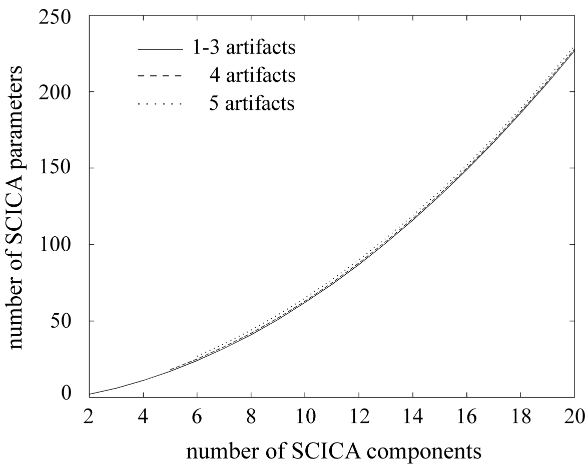

Fig. 6.10: Relationship between number of SCICA components (including 1-5 artifacts) and total number of parameters. The number of parameters is derived from the formulae given in Tables 6.3 and 6.4 in chapter 6.5.

The simplex algorithm is not very efficient in terms of the number of iterations (Press et al., 1992). For 5 SCICA components several hundred passes through the algorithm are needed until convergence, applying a fractional convergence tolerance of $1.e^{-4}$ (cf. chapter 6.1.1, step 7). For 20 SCICA components more than 10000 iterations may be required. These numbers hold for both the cumulant and the log-likelihood contrast. The actual number of iterations, however, strongly depends on the current random initialization. The number of iterations may be reduced by increasing the convergence tolerance.

Thus, calculating 5 SCICA components from a data segment of 2000 samples using the faster cumulant contrast and a fractional convergence tolerance of $1.e^{-4}$ amounts to a total duration of less than 2 seconds while deriving 20 SCICA components from an epoch of 4000 samples, again applying the cumulant contrast, may already require 20 minutes.

Summarizing, it is concluded that a total of 15 to 20 SCICA components should not be exceeded in the present SCICA implementation to allow for a computationally reasonable decomposition. Moreover, the length of one decomposed data segment should be reduced to the minimum necessary. In the ICA literature 10-s segments are recommended (Jung et al., 2000a). Experience suggests, however, that shorter epochs may be sufficient for SCICA due to the spatial constraint.

6.4. Discussion

The presented examples demonstrate that spatial filters derived from SCICA can remove artifacts from spontaneous EEG/MEG recordings without apparent distortion of brain activity (Figs. 6.3-6.8) irrespective of the spatial correlation between artifact and brain signal subspace (Fig. 6.2). A small quantified distortion in the simulated data was shown to be in the order of magnitude of recording noise and was, thus, considered to be irrelevant. Both the fourth-order auto-cumulant contrast and the log-likelihood contrast performed equally well. The cumulant contrast, however, is considerably faster than the log-likelihood objective function.

In order to allow for a computationally efficient estimation of the parameters by the simplex algorithm a total of 15 to 20 SCICA components may not be exceeded. Therefore, the dimension of the data segment which is equal to the number of SCICA components is always restricted to the minimum necessary. The dimension is the only parameter that has to be set for SCICA while the remaining parameters may be kept constant. The dimension can be estimated automatically. For spontaneous EEG/MEG, we have found it appropriate to reduce the dimensionality of the data to the number of eigenvectors each explaining at least 1 % of the total data variance. Our examples show that the number of eigenvectors estimated by the 1 % rule usually remains below the limit of 15/20. For raw EEG containing evoked activity of low SNR the 1 % rule may have to be lowered. Theoretically, there is no restriction to apply spatial filtering with SCICA to continuous event-related data. Thus, artifact correction with SCICA is equally applicable to spontaneous and event-related continuous recordings.

Before the SCICA decomposition is calculated it has to be checked whether the predefined artifacts are part of the current data segment. If a particular predefined artifact is present, its subspace correlation with the data segment is in general high. In the current implementation of SCICA a criterion based on the subspace correlation is employed successfully to exclude predefined artifact topographies automatically from artifact correction of a particular segment. Using automatic exclusion of predefined artifact topographies and the above 1 % rule, SCICA is sufficiently automated to be applied in combination with automated processing such as averaging or mean FFT analysis.

As the SCICA decomposition is an iterative process it is not as fast as the preselection approach. The current implementation in C requires a few seconds to several minutes on an Intel Pentium III-800 MHz processor depending on the number of SCICA components and time samples even if the faster cumulant contrast is applied. Currently, this prohibits an artifact correction during rapid clinical review of spontaneous EEG/MEG recordings. Artifacts may, however, be corrected in a separate step before any further analysis.

6.5. Appendix

In the Appendix specific algorithmic details of the SCICA decomposition referred to in section 6.1.1 are detailed.

6.5.1. A1: Expanding an orthonormal basis

Let $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_n) \in \Re^{l \times n}$, $1 \le n < l$ be a matrix with orthonormal columns. The following algorithm may be used to construct a matrix of orthonormal columns $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_l) \in \Re^{l \times l}$ with $\mathbf{W}(:,1:n) = \mathbf{V}$.

1. Set $\mathbf{W} = \mathbf{I}$ with identity matrix $\mathbf{I} \in \Re^{l \times l}$.
2. For $i = 1 \ldots n$: exchange vectors $\mathbf{w}_i$ and $\mathbf{w}_j$ if $\left|\mathbf{w}_j^{\mathrm{T}} \mathbf{v}_i\right| = \max$ for $j = i \ldots l$.
3. Set $\mathbf{W}(:,1:n) = \mathbf{V}$. Note that due to step 2 still $\operatorname{rank}(\mathbf{W}) = l$ which is a prerequisite for the modified Gram-Schmidt algorithm.
4. Apply the modified Gram-Schmidt algorithm (Meyer, 2000) to **W**:
   For $i = 2 \ldots l$:
   - For $j = i \ldots l$: $\mathbf{w}_j = \mathbf{w}_j - \left(\mathbf{w}_{i-1}^{\mathrm{T}} \mathbf{w}_j\right) \mathbf{w}_{i-1}$
   - $\mathbf{w}_i = \mathbf{w}_i \big/ \left\|\mathbf{w}_i\right\|$

Note that the orthonormal topographies $\mathbf{w}_2, \ldots, \mathbf{w}_n$ of step 3 are not changed by step 4.

6.5.2. A2: Perturbation of artifact topographies

The following algorithm is used to derive the perturbed artifact topographies $\overline{\mathbf{A}}^{(k)} = \left(\overline{\mathbf{a}}_1^{(k)},\ldots,\overline{\mathbf{a}}_n^{(k)}\right) \in \Re^{l \times n}$ from the original artifact topographies $\overline{\mathbf{A}} = \left(\overline{\mathbf{a}}_1,\ldots,\overline{\mathbf{a}}_n\right) \in \Re^{l \times n}$ in the k^{th} iteration.

1. For $i = 1\ldots n$: $\overline{\mathbf{a}}_i = \dfrac{\overline{\mathbf{a}}_i}{\left\|\overline{\mathbf{a}}_i\right\|}$.
2. Find a matrix $\mathbf{U}_i = \left(\mathbf{u}_{i,1},\ldots,\mathbf{u}_{i,l}\right) \in \Re^{l \times l}$ with orthonormal columns and $\mathbf{u}_{i,1} = \overline{\mathbf{a}}_i$ for $i = 1\ldots n$ referring, for instance, to the algorithm described in Appendix A1.
3. For $i = 1\ldots n$: $\overline{\mathbf{a}}_i^{(k)} = \overline{\mathbf{a}}_i + \tan\alpha_i^{(k)} \cdot \mathbf{v}_i^{(k)}$.

   The vector $\mathbf{v}_i^{(k)}$ is defined recursively according to $\mathbf{v}_i^{(k)} = \overline{\mathbf{v}}_{l-1}$

   $j = 1$: $\quad \overline{\mathbf{v}}_j = \mathbf{u}_{i,2}$

   for $j = 2\ldots(l-1)$: $\quad \overline{\mathbf{v}}_j = \cos\beta_{i,j-1}^{(k)} \cdot \overline{\mathbf{v}}_{j-1} + \sin\beta_{i,j-1}^{(k)} \cdot \mathbf{u}_{i,j+1}$

In Fig. 6.11 the procedure is illustrated for one artifact (n=1) in three dimensions (l=3). The angles α and β are randomly initialized and optimized in each iteration. For an overview of the number of angles and their initialization and optimization intervals refer to Table 6.3. The angle $\alpha_i^{(k)}$ is constrained to [-1.5°, 1.5°] in order to guarantee that $\left|\sphericalangle\left(\overline{\mathbf{a}}_i, \overline{\mathbf{a}}_i^{(k)}\right)\right| \leq 1.5°$ for $i = 1\ldots n$.

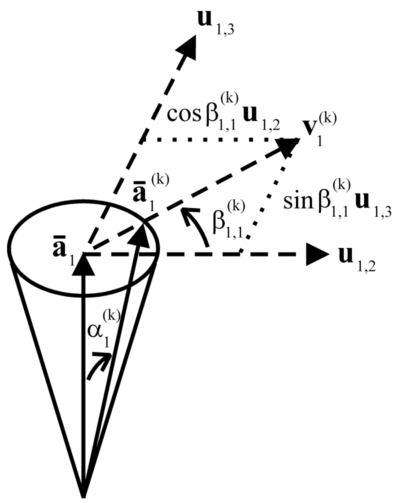

Fig. 6.11: The perturbed artifact topography $\overline{\mathbf{a}}_1^{(k)} \in \Re^3$ is derived from $\overline{\mathbf{a}}_1 \in \Re^3$ in the k^{th} iteration as described in Appendix A2. All vectors on the rim of the cone share the same angle $\alpha_1^{(k)}$ with $\overline{\mathbf{a}}_1$. The direction of the particular vector $\overline{\mathbf{a}}_1^{(k)}$ is determined by $\mathbf{v}_1^{(k)}$ i.e. $\beta_{1,1}^{(k)}$.

Table 6.3: Parameters for perturbation of n artifact topographies in $\Re^l$ as described in Appendix A2.

	number	random initialization in interval	optimization constrained to initial interval
α_i for $i=1\dots n$	1	[-1.5°,1.5°]	yes
β_i for $i=1\dots n$	l-2	[0°,180°]	no
$\sum$	$n(l$-1)		

6.5.3. A3: Linear combination matrix

Let $\mathbf{R}^{(k)} = \left(\mathbf{r}_1^{(k)},\dots,\mathbf{r}_n^{(k)},\mathbf{r}_{n+1}^{(k)},\dots,\mathbf{r}_l^{(k)}\right) \in \Re^{l\times l}$ be the linear combination matrix in the k^{th} iteration with $\mathbf{r}_i^{(k)} = \mathbf{e}_i$ for $i=1\dots n$, $1 \le n < l$. The vector $\mathbf{e}_i \in \Re^l$ is the i^{th} canonical vector. In order to construct the last l-n column vectors the following algorithm is used:

$$\text{For } i=(n+1)\dots l:\ \mathbf{r}_i^{(k)} = \begin{cases} \mathbf{u}_i^{(k)} + \tan\alpha_i^{(k)} \cdot \overline{\mathbf{r}}_i^{(k)} & \alpha_i^{(k)} \neq \dfrac{\pi}{2} \\ \overline{\mathbf{r}}_i^{(k)} & \alpha_i^{(k)} = \dfrac{\pi}{2} \end{cases}.$$

The vectors $\mathbf{r}_i^{(k)}$ and $\mathbf{u}_i^{(k)}$ enclose an angle of $\alpha_i^{(k)}$. The unit vector $\mathbf{u}_i^{(k)}$ lies in one of the planes spanned by $\overline{\mathbf{r}}_1^{(k)},\dots,\overline{\mathbf{r}}_{i-1}^{(k)}$. The vectors $\overline{\mathbf{r}}_1^{(k)},\dots,\overline{\mathbf{r}}_{i-1}^{(k)}$ form an orthonormal basis of the already determined column vectors $\mathbf{r}_1^{(k)},\dots,\mathbf{r}_{i-1}^{(k)}$ with $\overline{\mathbf{r}}_j^{(k)} = \mathbf{r}_j^{(k)}$ for $j=1\dots n$. The particular direction of $\mathbf{r}_i^{(k)}$ away from $\mathbf{u}_i^{(k)}$ is determined by unit vector $\overline{\mathbf{r}}_i^{(k)}$ that is perpendicular to $\overline{\mathbf{r}}_1^{(k)},\dots,\overline{\mathbf{r}}_{i-1}^{(k)}$. Consequently, the vectors $\overline{\mathbf{r}}_1^{(k)},\dots,\overline{\mathbf{r}}_i^{(k)}$ also form an orthonormal basis of $\operatorname{span}(\mathbf{r}_1^{(k)},\dots,\mathbf{r}_i^{(k)})$.

In Fig. 6.12 the algorithm is illustrated for one artifact (n=1) and two signal topographies (l=3). In the next two subsections the calculation of $\overline{\mathbf{r}}_1^{(k)},\dots,\overline{\mathbf{r}}_i^{(k)}$ and $\mathbf{u}_i^{(k)}$ is described.

Calculation of $\overline{\mathbf{r}}_1^{(k)},\dots,\overline{\mathbf{r}}_i^{(k)}$

The following recursive algorithm accumulates the orthonormal basis vectors $\overline{\mathbf{r}}_1^{(k)},\dots,\overline{\mathbf{r}}_i^{(k)}$ of $\operatorname{span}(\mathbf{r}_1^{(k)},\dots,\mathbf{r}_i^{(k)})$ in the first i columns of the orthogonal matrix $\overline{\mathbf{R}}_i^{(k)} \in \Re^{l\times l}$:

$$\text{for } i=(n+1)\dots(l-1): \qquad \overline{\mathbf{R}}_i^{(k)} = \overline{\mathbf{R}}_{i-1}^{(k)} \prod_{j=(i+1)\dots l} \mathbf{G}\left(i,j,\beta_{i,j}^{(k)}\right)$$

$$i=l: \qquad \overline{\mathbf{R}}_i^{(k)} = \overline{\mathbf{R}}_{i-1}^{(k)}$$

with $\overline{\mathbf{R}}_n^{(k)} = \mathbf{I} \in \Re^{l\times l}$ and the Givens rotation matrix

$$\mathbf{G}(i,j,\beta)=\begin{array}{cc}
\begin{bmatrix}1 & \cdots & 0 & \cdots & 0 & \cdots & 0\\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos\beta & \cdots & -\sin\beta & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots\\ 0 & \cdots & \sin\beta & \cdots & \cos\beta & \cdots & 0\\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1\end{bmatrix} & \begin{matrix} \\ \\ i \\ \\ j \\ \\ \\ \end{matrix} \\
\begin{matrix} & & i & & j & & \end{matrix} &
\end{array}$$

that rotates the plane spanned by the i^{th} and j^{th} column vector of $\overline{\mathbf{R}}_{i-1}^{(k)}$ by an angle of β (Golub and van Loan, 1996). Note that the first (i-1) columns of $\overline{\mathbf{R}}_{i}^{(k)}$ and $\overline{\mathbf{R}}_{i-1}^{(k)}$ are equal as Givens rotations are applied to columns of $\overline{\mathbf{R}}_{i-1}^{(k)}$ with an index larger than i-1 only. The i^{th} column vector of $\overline{\mathbf{R}}_{i}^{(k)}$ is determined in the i^{th} iteration.

Calculation of $\mathbf{u}_i^{(k)}$

The vector $\mathbf{u}_i^{(k)}$ lies in one of the planes spanned by $\overline{\mathbf{r}}_1^{(k)},\ldots,\overline{\mathbf{r}}_{i-1}^{(k)}$:

$i = 2$: $\quad \mathbf{u}_i^{(k)} = \overline{\mathbf{r}}_1^{(k)} = \mathbf{e}_1$

$i > 2$: $\quad \mathbf{u}_i^{(k)} = \cos\delta_i^{(k)}\overline{\mathbf{r}}_s^{(k)} + \sin\delta_i^{(k)}\overline{\mathbf{r}}_t^{(k)}, \quad 1 \le s < t \le i-1$.

The $\binom{i-1}{2} = \frac{(i-1)(i-2)}{2}$ possible planes, i.e. combinations of s and t, are coded in the angle $\delta_i^{(k)}$. For a given angle $\delta_i^{(k)}$ with $(w-1)\cdot 180° \le \delta_i^{(k)} < w \cdot 180°, 1 \le w \le \binom{i-1}{2}$, the parameters s and t are chosen as the w^{th} combination of s and t.

The angles α, β and δ are randomly initialized and optimized in each iteration. For an overview of the number of angles and their initialization and optimization intervals refer to Table 6.4. The angle $\alpha_i^{(k)}$ is constrained to the interval [10°, 170°] in order to guarantee that $10° \le \left(\text{span}\left\{\mathbf{r}_1^{(k)},\ldots,\mathbf{r}_{i-1}^{(k)}\right\},\mathbf{r}_i^{(k)}\right) \le 170°$ for $i = (n+1)\ldots l$.

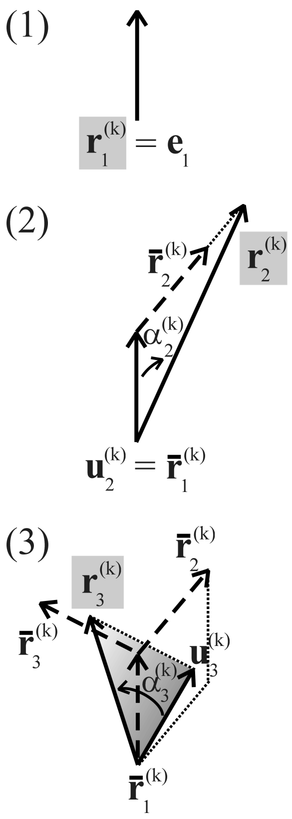

Fig. 6.12: Constructing the three column vectors $\mathbf{r}_1^{(k)}, \mathbf{r}_2^{(k)}, \mathbf{r}_3^{(k)}$ of the linear combination matrix $\mathbf{R}^{(k)}$ for one artifact (n=1) and two signal topographies in $\Re^3$ (l=3) in the k^{th} iteration. The notation corresponds to the algorithm described in Appendix A3.

(1) The first column vector $\mathbf{r}_1^{(k)}$ is the first canonical vector $\mathbf{e}_1 = (1,0,0)^{\mathrm{T}}$ in $\Re^3$. Thus, the artifact topography in the first column of $\overline{\mathbf{C}}_{\mathrm{o}}^{(k)}$ remains unchanged by linear combination.

(2) The second column vector $\mathbf{r}_2^{(k)}$ is $\alpha_2^{(k)}$ away from $\mathbf{u}_2^{(k)} = \overline{\mathbf{r}}_1^{(k)} = \mathbf{r}_1^{(k)} = \mathbf{e}_1$. There is an infinite number of vectors sharing the same angle $\alpha_2^{(k)}$ with $\mathbf{u}_2^{(k)}$. The direction of the particular vector $\mathbf{r}_2^{(k)}$ is determined by $\overline{\mathbf{r}}_2^{(k)}$ that is perpendicular to $\overline{\mathbf{r}}_1^{(k)}$. The vectors $\overline{\mathbf{r}}_1^{(k)}$, $\overline{\mathbf{r}}_2^{(k)}$ form an orthonormal basis of the vector space spanned by $\mathbf{r}_1^{(k)}$, $\mathbf{r}_2^{(k)}$.

(3) The third column vector $\mathbf{r}_3^{(k)}$ is $\alpha_3^{(k)}$ away from $\mathbf{u}_3^{(k)}$ in direction of $\overline{\mathbf{r}}_3^{(k)}$. The vector $\mathbf{u}_3^{(k)}$ lies in the plane spanned by $\overline{\mathbf{r}}_1^{(k)}$ and $\overline{\mathbf{r}}_2^{(k)}$. The vector $\overline{\mathbf{r}}_3^{(k)}$ is perpendicular to $\overline{\mathbf{r}}_1^{(k)}$, $\overline{\mathbf{r}}_2^{(k)}$. The vectors $\overline{\mathbf{r}}_1^{(k)}, \overline{\mathbf{r}}_2^{(k)}, \overline{\mathbf{r}}_3^{(k)}$ form an orthonormal basis of the vector space spanned by $\mathbf{r}_1^{(k)}, \mathbf{r}_2^{(k)}, \mathbf{r}_3^{(k)}$.

Table 6.4: Parameters for linear combination of l-n signal topographies in $\Re^l$ as described in Appendix A3. For i=2 (i.e. n=1) there is no δ_i. In this case the total number (Σ) is reduced by 1!

	number	random initialization in interval	optimization constrained to initial interval
α_i for $i = (n+1)\ldots l$	1	[10°,170°]	yes
β_i for $i = (n+1)\ldots l$	l-i	[0°,180°]	no
δ_i for $i = (n+1)\ldots l$	1	$\left[0°, \left(\frac{i-1}{2}\right)180°\right]$	yes
$\sum$	$\frac{(l-n)(l-n+3)}{2}$		

6.5.4. A4: Whitening matrix

Let $\mathbf{S}^{(k)} = \left(\overline{\mathbf{C}}^{(k)}\right)^{-1}\overline{\mathbf{D}} \in \Re^{l \times t}$ be a matrix whose rows contain the waveforms corresponding to data matrix $\overline{\mathbf{D}} \in \Re^{l \times t}$ and topography matrix $\overline{\mathbf{C}}^{(k)} \in \Re^{l \times l}$ in the k^{th} iteration. Let $\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}}$ be the SVD of $\frac{1}{t}\mathbf{S}^{(k)}\left(\mathbf{S}^{(k)}\right)^{\mathrm{T}}$. Then $\overline{\mathbf{W}}^{(k)} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{V}^{\mathrm{T}} \in \Re^{l \times l}$ is a whitening matrix fulfilling condition 1, i.e. $\mathrm{E}\left\{\overline{\mathbf{W}}^{(k)}\mathbf{S}^{(k)}\left(\overline{\mathbf{W}}^{(k)}\mathbf{S}^{(k)}\right)^{\mathrm{T}}\right\}=\mathbf{I}$. Note that $\overline{\mathbf{W}}^{(k)}$ can only be determined if $\frac{1}{t}\mathbf{S}^{(k)}\left(\mathbf{S}^{(k)}\right)^{\mathrm{T}}$ has full column rank. By a series of $x=\sum_{i=1}^{n}(l-i)$ suitable Givens rotations, $\overline{\mathbf{W}}^{(k)}$ may be transformed into $\mathbf{W}^{(k)}=\mathbf{G}_x\ldots\mathbf{G}_1\overline{\mathbf{W}}^{(k)}\in\Re^{l\times l}$ with

$$\mathbf{W}^{(k)}=\begin{bmatrix} w_{1,1} & \cdots & w_{1,n} & w_{1,n+1} & \cdots & w_{1,l} \\ 0 & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & w_{n,n} & w_{n,n+1} & \cdots & w_{n,l} \\ \vdots & & 0 & w_{n+1,n+1} & \cdots & w_{n+1,l} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & w_{l,n+1} & \cdots & w_{l,l} \end{bmatrix}$$

having zeros below the diagonal in the first n columns (Golub and van Loan, 1996). With the inverse $\left(\mathbf{W}^{(k)}\right)^{-1}$ having zeros below the diagonal in the first n columns, too, the first n column vectors $\overline{\mathbf{c}}_{\mathrm{W},1}^{(k)},\ldots,\overline{\mathbf{c}}_{\mathrm{W},n}^{(k)}$ of $\overline{\mathbf{C}}_{\mathrm{W}}^{(k)}=\overline{\mathbf{C}}^{(k)}\left(\mathbf{W}^{(k)}\right)^{-1}$ are a linear combination of the first n column vectors $\overline{\mathbf{c}}_1^{(k)},\ldots,\overline{\mathbf{c}}_n^{(k)}$ of $\overline{\mathbf{C}}^{(k)}$ only. Thus, $\mathbf{W}^{(k)}$ fulfills conditions 1 and 2:

- $\mathrm{E}\left\{\mathbf{W}^{(k)}\mathbf{S}^{(k)}\left(\mathbf{W}^{(k)}\mathbf{S}^{(k)}\right)^{\mathrm{T}}\right\}=\mathbf{I}$
- $\mathrm{span}\left(\overline{\mathbf{c}}_{\mathrm{W},1}^{(k)},\ldots,\overline{\mathbf{c}}_{\mathrm{W},n}^{(k)}\right)=\mathrm{span}\left(\overline{\mathbf{c}}_1^{(k)},\ldots,\overline{\mathbf{c}}_n^{(k)}\right)$.

6.5.5. A5: Orthogonal log-likelihood contrast

From the normalized (i.e. divided by t) natural logarithm of the likelihood L of the source separation model $\overline{\mathbf{D}}=\overline{\mathbf{C}}_{\mathrm{W}}^{(k)}\mathbf{S}_{\mathrm{W}}^{(k)}\in\Re^{l\times t}$ in the k^{th} iteration with $\overline{\mathbf{C}}_{\mathrm{W}}^{(k)}\in\Re^{l\times l}$ and $\mathbf{S}_{\mathrm{W}}^{(k)}=\left(\mathbf{s}_{\mathrm{W},1}^{(k)},\ldots,\mathbf{s}_{\mathrm{W},l}^{(k)}\right)^{\mathrm{T}}\in\Re^{l\times t}$ the log-likelihood contrast ϕ_{L} can be derived (Pham et al., 1992; cf. chapter 3.4.4):

$$\phi_{\mathrm{L}}=\frac{1}{t}\ln\mathrm{L}=\frac{1}{t}\sum_{j=1}^{t}\sum_{i=1}^{l}\ln\mathrm{p}_i\left(\mathbf{s}_{\mathrm{W},i}^{(k)}(j)\right)+\ln\left|\det\left(\overline{\mathbf{C}}_{\mathrm{W}}^{(k)-1}\right)\right|=\max$$

where $\mathrm{p}_i\left(\mathbf{s}_{\mathrm{W},i}^{(k)}(j)\right)$ is the probability density function of the i^{th} whitened waveform.

Under the whiteness constraint $\mathrm{E}\left\{\mathbf{S}_{\mathrm{W}}^{(k)}\left(\mathbf{S}_{\mathrm{W}}^{(k)}\right)^{\mathrm{T}}\right\}=\overline{\mathbf{C}}_{\mathrm{W}}^{(k)-1}\mathrm{E}\left\{\overline{\mathbf{D}}\,\overline{\mathbf{D}}^{\mathrm{T}}\right\}\overline{\mathbf{C}}_{\mathrm{W}}^{(k)-\mathrm{T}}=\mathbf{I}$. With $\det\mathbf{I}=1=\left(\det\overline{\mathbf{C}}_{\mathrm{W}}^{(k)-1}\right)\det\mathrm{E}\left\{\overline{\mathbf{D}}\,\overline{\mathbf{D}}^{\mathrm{T}}\right\}\det\left(\overline{\mathbf{C}}_{\mathrm{W}}^{(k)-\mathrm{T}}\right)$ and $\det\mathrm{E}\left\{\overline{\mathbf{D}}\,\overline{\mathbf{D}}^{\mathrm{T}}\right\}=\mathrm{const}$ it follows $\det\overline{\mathbf{C}}_{\mathrm{W}}^{(k)-1}$ is also constant (Hyvärinen and Oja, 2000). This leads to the following orthogonal contrast function:

$$\phi_{\mathrm{L}}^{\mathrm{o}}=\frac{1}{t}\sum_{j=1}^{t}\sum_{i=1}^{l}\ln\mathrm{p}_i\left(\mathbf{s}_{\mathrm{W},i}^{(k)}(j)\right)=\max\quad\text{respectively}\quad\phi_{\mathrm{L}}^{\mathrm{o}}=-\frac{1}{t}\sum_{j=1}^{t}\sum_{i=1}^{l}\ln\mathrm{p}_i\left(\mathbf{s}_{\mathrm{W},i}^{(k)}(j)\right)=\min.$$

7. Advantages of SCICA over ICA

In this chapter advantages of SCICA over the closely related ICA are demonstrated in the context of artifact correction. The major drawback of ICA is that artifact components have to be identified manually in the absence of a suitable automated detection approach. In SCICA this problem does not occur because artifact topographies are incorporated as prior knowledge into the decomposition. Being a conceivable alternative to SCICA, a simple automated detection approach has been developed that identifies artifact ICA components after the decomposition by correlating each ICA topography with the predefined artifact subspace. In chapter 7.1 this approach is applied and compared with SCICA. Although automated detection by topography is successful in many cases, the examples demonstrate that the approach fails when eye movement artifacts are decomposed into (multiple) ICA components exhibiting only a low correlation with the predefined artifact subspace. Therefore, incorporating artifact topographies into the decomposition seems to be superior to making use of their prior knowledge after the decomposition.

In chapter 7.2 SCICA and ICA are compared in the case of temporally dependent artifact and brain signal activity. In the first example it is shown that SCICA is clearly superior to ICA if brain and artifact activity occur simultaneously for some time and subepochs of brain signal activity without artifact exist. While these periods of isolated brain activity are sufficient for SCICA, ICA needs at least one further artifact in a different context to express its functional independence from the brain activity. The second example uses simulated data to illustrate that neither ICA nor SCICA are suitable if artifact and cerebral activity occur (nearly) synchronously only as, for instance, in averaged evoked data. SCICA can separate temporally dependent components, however, if their topographies are known in advance.

ICA analysis was performed with Matlab implementations either of the JADE or the extended infomax algorithm (see chapter 3.4.5). Due to their close relationship the JADE algorithm was compared to SCICA with the fourth-order auto-cumulant contrast, and the extended infomax algorithm was contrasted to SCICA using the log-likelihood objective function. To be as close as possible to our implementation the kurtosis switching rule in the extended ICA implementation was replaced by the stability switching rule (Lee et al., 1999).

Before artifact correction the data dimension was reduced (cf. step 1 of chapter 6.1.1). For real data, only eigenvectors explaining at least 1 % of the total data variance were retained (1 % rule). For simulated data, the number of eigenvectors was set to the rank of the simulated epoch (rank rule).

7.1. SCICA and ICA with automated artifact detection by topography

In the following two examples artifact correction by SCICA is compared to artifact correction by ICA with an automated artifact detection by topography. ICA components are considered as artifacts if their topographies exceed a certain correlation, e.g. 90-95 %, with the predefined SCICA artifact subspace. The examples demonstrate why an automated detection of ICA components by topography is limited.

7.1.1. Example 1: Removing blinks and eye movements

In the first example artifact correction by SCICA using the log-likelihood contrast is compared to extended infomax with automated detection of artifact components. Before artifact correction the dimensionality of the data was reduced to 7 eigenvectors (1 % rule).

In the upper left panel of Fig. 7.1 the original EEG epoch (10 s, 1-70 Hz, referenced to the average of F3 and F4) containing horizontal eye movements and blinks is depicted. In the upper right panel the SCICA corrected segment and the SCICA waveforms corresponding to horizontal eye movement and blink topography are shown. The eye artifacts are removed without recognizable distortion of the delta activity at parieto-occipital electrodes. Artifact topographies were derived from averaged blinks and horizontal eye movements. In the lower left panel the 7 equally scaled ICA waveforms are displayed. Visual inspection of waveforms and topographies reveals that the blinks are clearly concentrated in component 3 whereas the horizontal eye movements are less clearly spread over components 2 and 6. Component 6 also contributes to a negative potential at posterior electrodes (note areas of negative potential are hatched in the maps). This explains the lower subspace correlation of 85.6 % with the SCICA artifact subspace in comparison to 95.5 % and 98.9 % for components 2 and 3. The correlation of the remaining ICA topographies with the SCICA artifact subspace is 79.3 % (ICA 1), 70.6 % (ICA 4), 67.5 % (ICA 5) and 40.6 % (ICA 7). In this example, the automated detection of artifact ICA components reveals only artifact components 2/3 assuming a reasonable threshold of 90-95 %. Removing these two components yields the segment shown in the lower right panel. Traces Fp2, F7, F8, F9 and F10 still contain a considerable amount of horizontal eye artifacts. After eliminating ICA component 6 as well, the eye artifacts are corrected. The distortion at posterior electrodes introduced by component 6 is below visibility in this case.

7.1.2. Example 2: Removing cardiac artifacts and eye movements

In the second example artifact correction by SCICA using the fourth-order cumulant-based contrast is compared to the JADE algorithm with an automated detection of artifact components. Before artifact correction the dimensionality of the data was reduced to 8 eigenvectors (1 % rule).

In the upper left panel of Fig. 7.2 the original EEG epoch (10 s, 1-70 Hz, referenced to the average of F3 and F4) recorded with eyes closed is depicted. The epoch contains cardiac artifacts and small vertical eye movements under the closed eye-lids. In the upper right panel the SCICA corrected segment and the SCICA waveforms corresponding to vertical eye movement and EKG topography are shown. The artifacts are removed without apparent distortion of the alpha rhythm. The vertical eye movement topography was derived from a single vertical eye movement. The EKG topography was determined from the averaged QRS-complex of several hundreds of cardiac cycles. In the lower left panel the 8 equally scaled ICA waveforms are displayed. Visual inspection of the waveforms shows that the cardiac artifacts are modeled by component 4 whereas the ocular artifacts are represented by component 6 and also to some extent by component 1. However, the automated detection of artifact ICA components reveals only component 4 with a subspace correlation of 98.8 % applying again a reasonable threshold of 90-95 %. As component 6 also contributes to a positive potential at right temporal and posterior electrodes, its subspace correlation with the predefined artifact subspace amounts merely to 69.8 %. Component 1 is slightly oblique and has, therefore, only a subspace correlation of 86 %. The correlation of the remaining ICA topographies with the SCICA artifact subspace is 55.3 % (ICA 2), 78.7 % (ICA 3), 56.1 % (ICA 5), 35.6 % (ICA 7), and 33.5 % (ICA 8). Removing the automatically detected component 4 yields the segment shown in the lower right panel. The ocular artifacts are not corrected. Applying ICA component 6 as well, the eye movements are eliminated solely in channels Fp1 and Fp2. The distortions introduced by component 6 are barely visible, though. Additionally employing ICA component 1, the eye artifacts are completely removed. However, part of the alpha activity is reduced, too, for example in the first 4 seconds in channels A1, A2, F9, F10, P9, and P10.

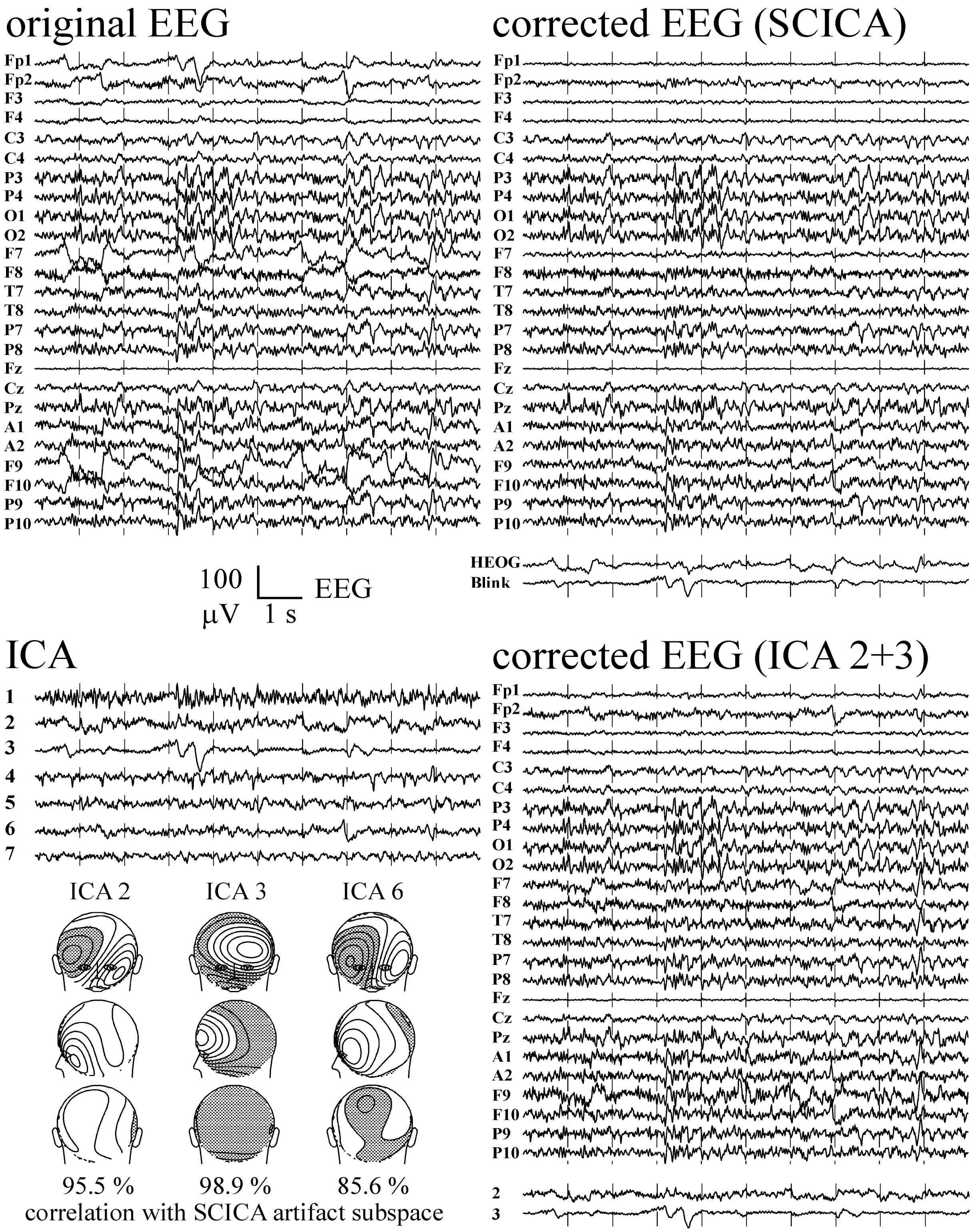

Fig. 7.1: Comparing SCICA and ICA with automated detection of artifact components by topography (cf. Example 1 chapter 7.1.1). Upper left: The original EEG segment (10 s, 1-70 Hz) superimposed by blinks and horizontal eye movements. Upper right: EEG segment after artifact removal by SCICA. The reconstructed waveforms corresponding to the predefined horizontal eye movement and blink topography are shown in separate traces at the bottom. Lower left: ICA waveforms and scalp maps for selected components representing eye artifacts. Below the maps the subspace correlation of the ICA topography with the SCICA artifact subspace is indicated. Applying a reasonable threshold of 90-95 % only ICA components 2 and 3 are automatically identified as artifacts. Lower right: Segment after removing the automatically detected ICA components 2 and 3. Artifact removal is incomplete.

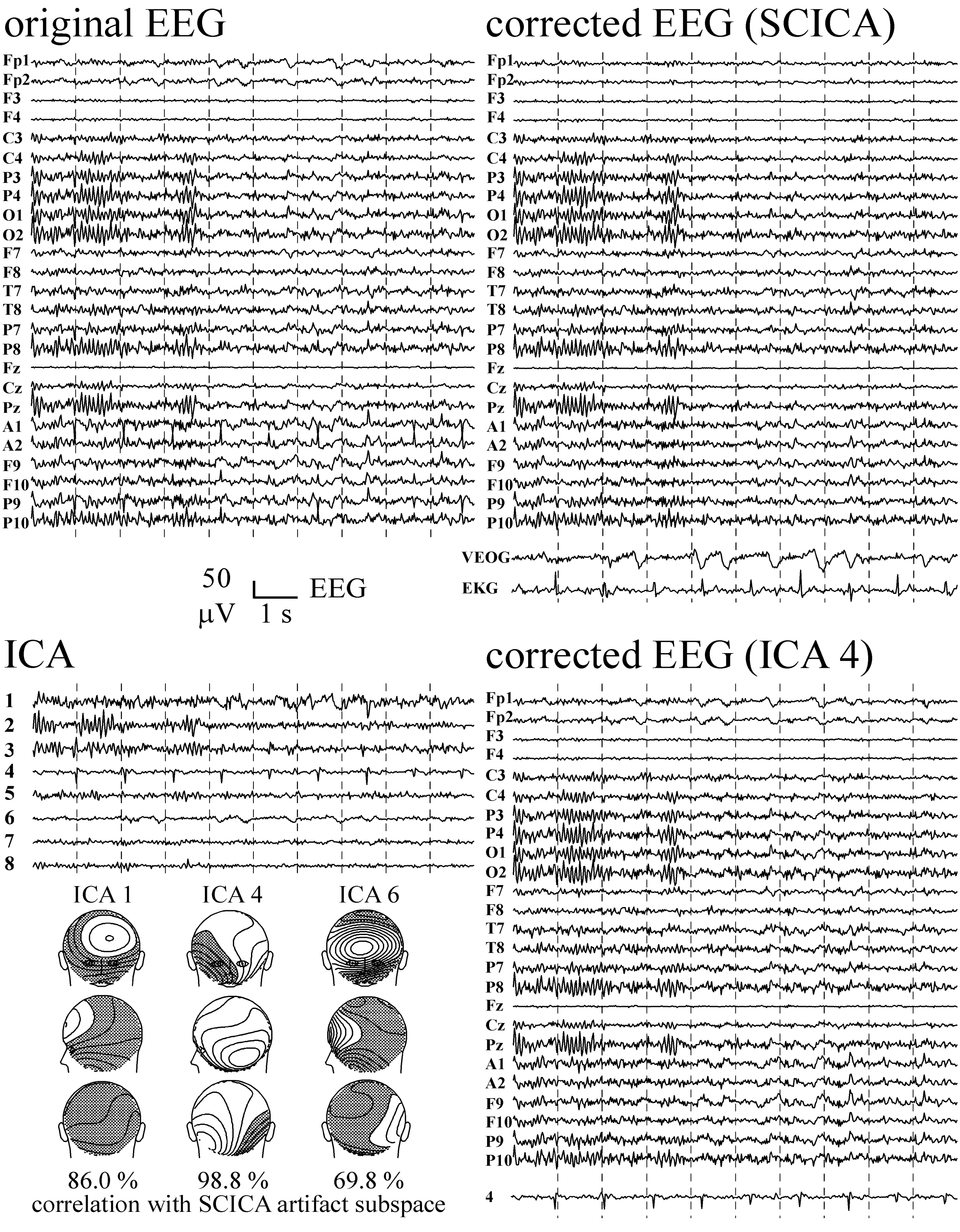

Fig. 7.2: Comparing SCICA and ICA with automated detection of artifact components by topography (cf. Example 2 chapter 7.1.2). Upper left: The original EEG segment (10 s, 1-70 Hz) superimposed by cardiac artifacts and small vertical eye movements. Upper right: EEG segment after artifact removal by SCICA. The reconstructed waveforms corresponding to the predefined vertical eye movement and EKG topography are shown in separate traces at the bottom. Lower left: ICA waveforms and scalp maps for selected components representing artifacts. Below the maps the subspace correlation of the ICA topography with the SCICA artifact subspace is indicated. Applying a reasonable threshold of 90-95 % only ICA component 4 is automatically identified as an artifact. Lower right: EEG segment after removing the automatically detected ICA component 4. The eye artifacts are not eliminated.

7.2. SCICA and ICA in the case of temporally dependent activity

In the following examples artifact correction by SCICA and ICA is compared in the case of temporally dependent artifact and brain activity. The examples demonstrate that incorporating prior knowledge about the artifact topographies makes SCICA superior to ICA in the case of temporally dependent activity.

7.2.1. Example 1: Subepochs of signal activity without artifact

In the first example, SCICA using the fourth-order cumulant-based contrast is compared to the JADE algorithm. Before artifact correction the dimensionality of the data was reduced to 8 eigenvectors (1 % rule).

In panel (A) of Fig. 7.3 the original EEG segment (10 s, 1-70 Hz, referenced to the average of F3 and F4) is shown. The epoch is contaminated by a single blink in the 4th second. The blink coincides temporally with one burst of a 3-Hz rhythmic activity most prominent at electrodes P3 and O1. Thus, artifact and brain signal activity are correlated in this time range. In panel (B) the 8 equally scaled SCICA waveforms are displayed. Waveform 1 corresponds to the predefined blink topography depicted below the SCICA waveforms. The blink topography was defined on an averaged blink. An overplot of the original (red) and the SCICA corrected (black) EEG is shown in panel (D). The blink is corrected without apparent distortion of the rhythmic discharges. The amplitude of the delta burst at P3 and O1 in the marked time range of the blink is more positive after SCICA correction (note positivity is shown downwards) due to the eliminated negative contribution of the blink at posterior electrodes. In panel (C) the 8 equally scaled ICA waveforms are displayed. Visual inspection of the waveforms shows that waveform 6 clearly depicts the blink activity. The corresponding topography, however, reveals that ICA erroneously combines the blink activity with the temporally dependent discharge at electrodes P3 and O1 that is part of the rhythmic delta activity. Eliminating ICA component 6 from the original EEG as shown in the overplot of original (red) and ICA corrected (black) EEG in panel (E) thus not only corrects the blink but also eliminates the temporally coinciding delta burst at electrodes P3 and O1 in the marked time range of overlap.

In panel (A) of Fig. 7.4 the same EEG as in Fig. 7.3 is shown 2 seconds later. The blink that was in the 4th second in the previous example is now in the 2nd second. In the last second another blink occurs. The additional blink appears in a functionally different context as it does not overlap with brain signal activity. Before artifact correction the dimensionality of the data was again reduced to 8 eigenvectors (1 % rule). In panel (B) the 8 equally scaled SCICA waveforms are displayed. Waveform 1 corresponds to the predefined blink topography depicted below the SCICA waveforms. In panel (C) the equally scaled ICA waveforms are shown. Waveform 3 clearly contains the blinks. Due to the additional blink, the corresponding topography 3 exclusively represents the blinks in this case. The overplot of original and SCICA/ICA corrected EEG is shown in panels (D) and (E). Using either method the blinks are removed without recognizable distortion of the rhythmic brain activity.

7.2.2. Example 2: No subepochs of signal activity without artifact

In the last example temporally dependent activity is simulated as it may, for instance, occur in averaged evoked segments. In panel (A) of Fig. 7.5 the dipole simulation using a 4-shell spherical head model is depicted. The waveforms corresponding to dipoles 1 (artifact) and 2 (signal) are temporally dependent. Signal waveforms 3 and 4 are clearly independent in time. In panel (B) the 2-s simulated data set is shown. The electrode configuration includes the standard 10-20 electrodes plus 3 infero-temporal electrodes on either side of the head and 4 eye electrodes. No noise was added to the simulated data. In panel (C) ICA and SCICA are compared. Before ICA or SCICA decomposition the dimensionality of the data was reduced to 4 eigenvectors (rank rule). Applying the JADE algorithm yields the waveforms and topographies shown in the left column. ICA component 2 mainly represents the temporally dependent activity of dipoles 1 and 2 while component 1 predominantly models the activity in the small time range where dipole 1 is active without dipole 2. ICA components 3 and 4 are quite similar to the simulated dipoles 3 and 4. Applying SCICA with the fourth-order cumulant-based contrast and topography 1 predefined yields the waveforms and topographies displayed in the middle column. Although the predefined SCICA topography 1 is correct, SCICA component 2 still erroneously combines the correlated activity of dipoles 1 and 2. Therefore, SCICA component 1 keeps representing the solo activity of dipole 1. SCICA components 3 and 4, on the other hand, are almost equivalent to the simulated dipoles 3 and 4. Finally, in the right column it is demonstrated that SCICA can separate the temporally dependent sources 1 and 2 provided that both their topographies are known in advance. In general, however, no prior information will be available about a signal topography.

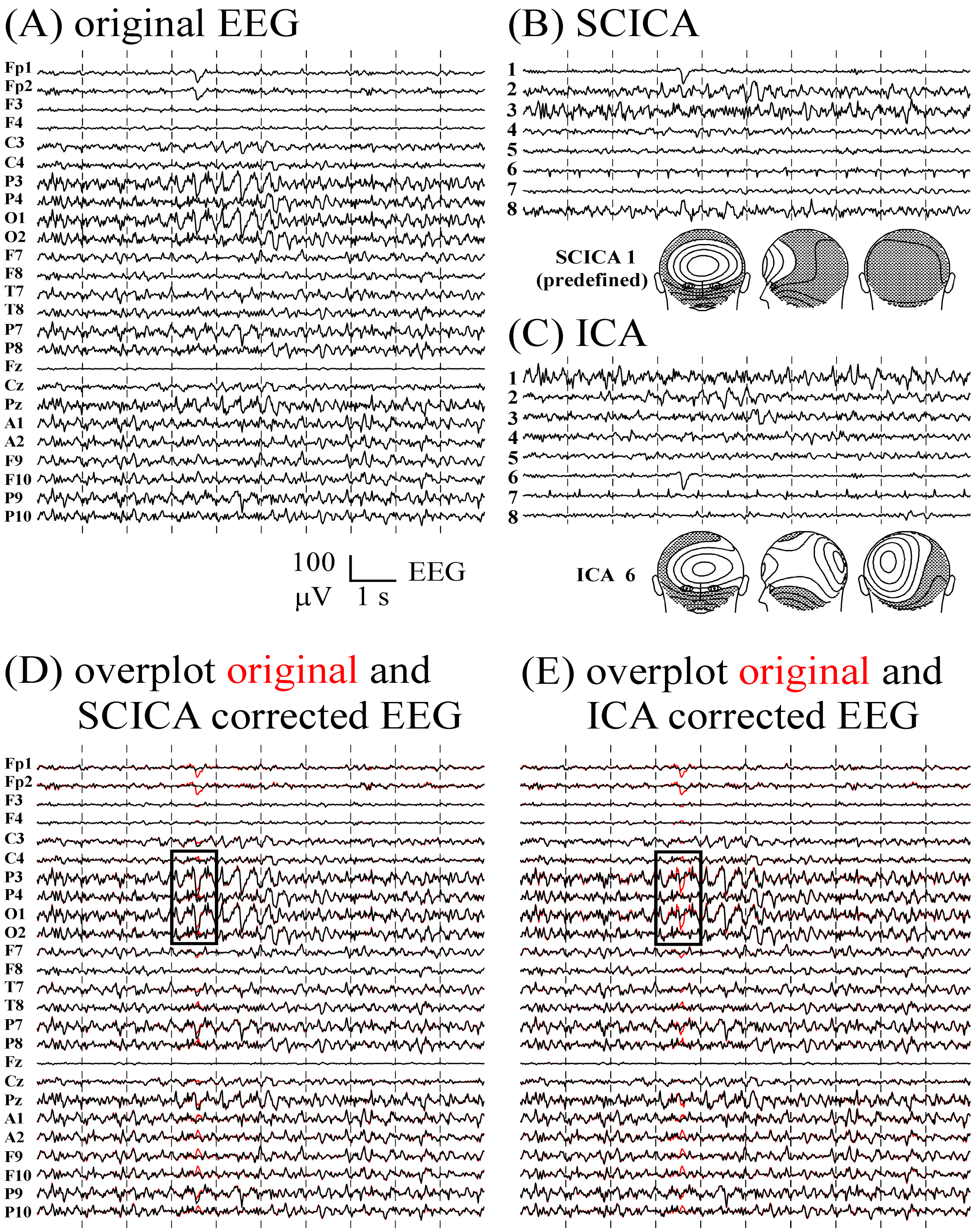

Fig. 7.3: Comparing SCICA and ICA in the case of temporally dependent brain and artifact activity (cf. Example 1 chapter 7.2.1). (A) The original EEG segment (10 s, 1-70 Hz) is superimposed by a blink that temporally coincides with one burst of the rhythmic brain activity at electrodes P3 and O1. (B) The 8 equally scaled SCICA waveforms and the predefined blink topography. (C) The 8 equally scaled ICA waveforms. Waveform 6 represents the blink activity. The corresponding topography erroneously combines blink and temporally dependent brain activity. (D) Overplot of original and SCICA corrected EEG segment. The brain activity at electrodes P3 and O1 in the marked time range is preserved. (E) Overplot of original and ICA corrected EEG segment. The brain activity at electrodes P3 and O1 in the marked time range is distorted.

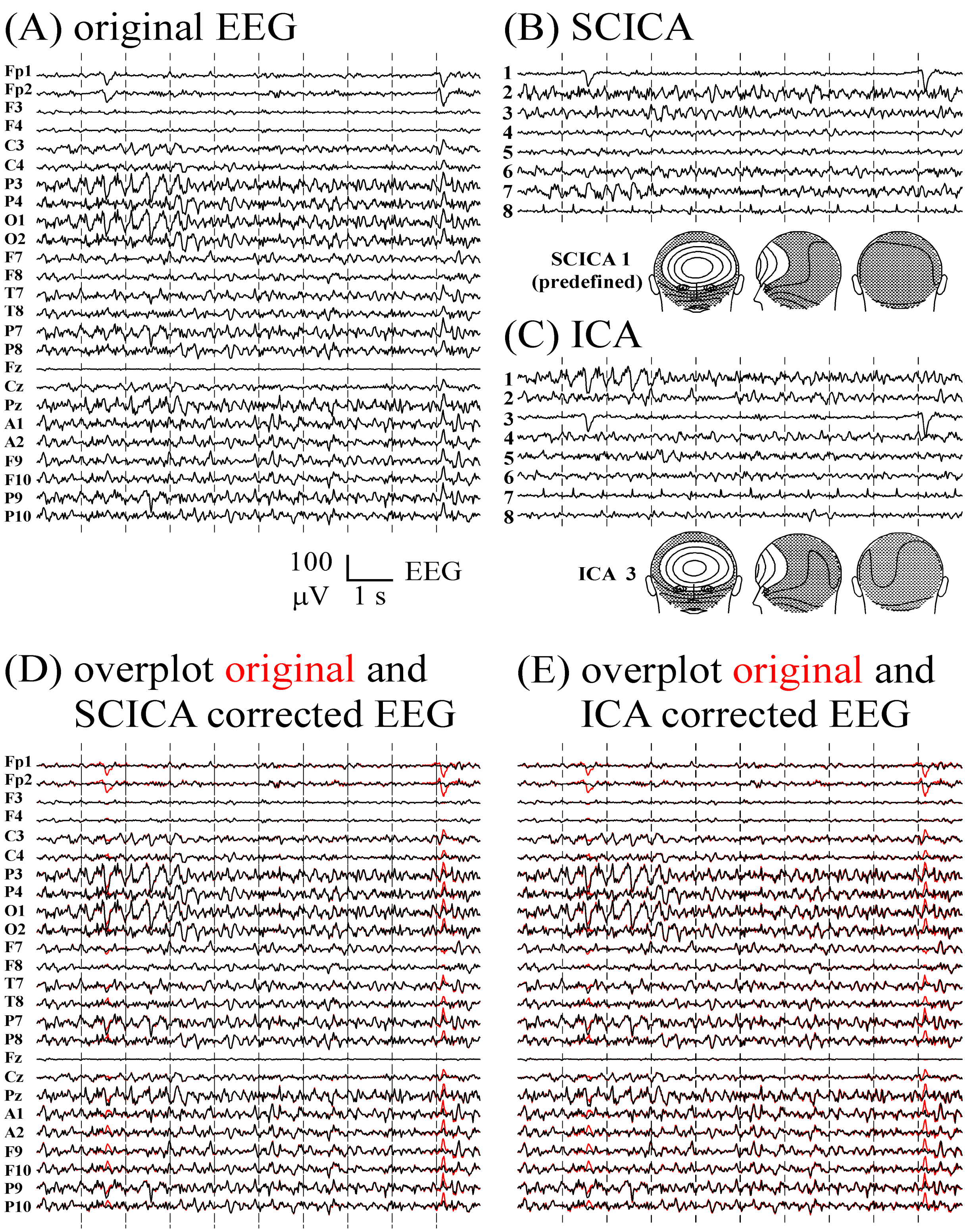

Fig. 7.4: Comparing SCICA and ICA in the case of temporally dependent brain and artifact activity (cf. Example 1 chapter 7.2.1). (A) The original EEG segment (10 s, 1-70 Hz) starting 2 s later than the epoch shown in Fig. 7.3 is still superimposed by the blink that temporally coincides with one burst of the rhythmic brain activity at electrodes P3 and O1 and an additional blink appearing in a different functional context. (B) The 8 equally scaled SCICA waveforms and the predefined blink topography. (C) The 8 equally scaled ICA waveforms. Waveform and topography 3 correctly represent the blinks. (D) Overplot of original and SCICA corrected EEG segment. (E) Overplot of original and ICA corrected EEG segment. Using either method the brain activity at electrodes P3 and O1 in the marked time range is preserved.

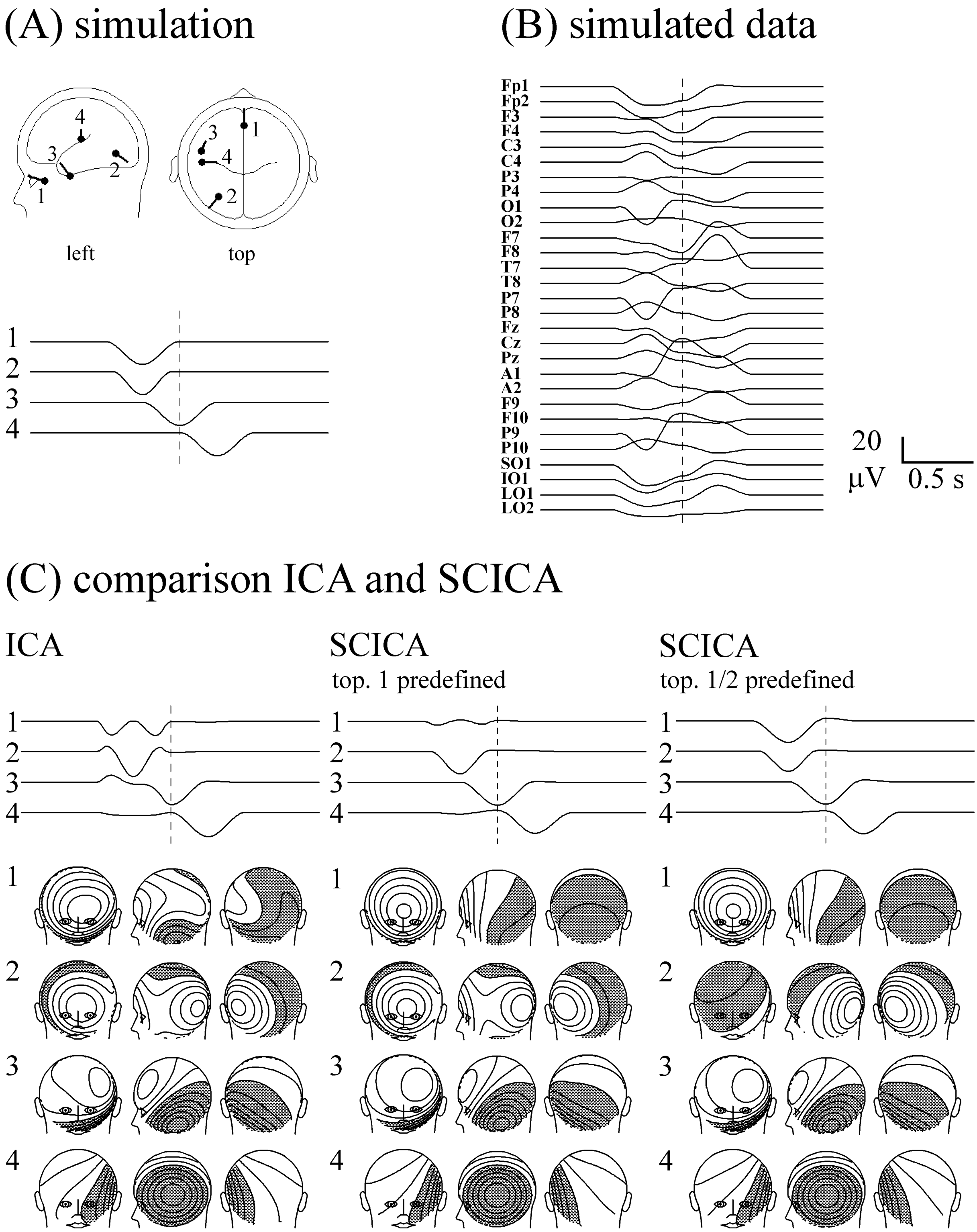

Fig. 7.5: Comparing SCICA and ICA in the case of simulated temporally dependent activity (cf. Example 2 chapter 7.2.2). (A) The dipole simulation. Waveforms 1 and 2 are temporally dependent. (B) The simulated data set (2 s). (C) ICA and SCICA decomposition of the simulated data. Only SCICA with topographies 1 and 2 predefined (right column) can recover the simulated waveforms and topographies completely.

8. Summary and Discussion

In this thesis two spatial filter approaches for artifact correction in continuous EEG and MEG recordings have been presented. The spatial filters are based on artifact and brain signal topographies comparable to the spatial filters derived from MSEC and ICA. Analogous to MSEC, artifact topographies are derived in advance from single or averaged artifacts of the same recording session. In order to estimate brain signal topographies two novel approaches have been introduced in this dissertation: preselection and SCICA.

In the preselection approach brain signal topographies are determined as the eigenvectors of an artifact-free subset of the data segment. The subset is obtained by excluding sample vectors that exceed a certain amplitude or correlation with the predefined artifact subspace. The quality of artifact correction with preselection depends crucially on the subjective choice of amplitude threshold, correlation threshold and number of eigenvectors. The parameters are hard to estimate automatically. Incorrect thresholds may result in severe distortion of brain activity or noise enhancement. The preselection approach is especially useful for artifact correction during review of spontaneous EEG/MEG recordings as it is sufficiently fast and as suitable parameter thresholds may be found empirically. The only drawback is that parameters may have to be adjusted manually from one epoch to another. The approach seems to be less appropriate to artifact removal in event-related continuous data. An optimal number of eigenvectors may not be found in this case due to the tradeoff between low SNR event-related signals and remaining low SNR artifacts in the 'artifact-free' subset. As parameters cannot be adjusted automatically, the preselection approach is also not suitable for artifact correction in combination with automated processing such as averaging or mean FFT analysis.

The novel concept of SCICA offers an alternative way of modeling the brain signal subspace. The basic idea of SCICA is to incorporate the prior knowledge about artifact topographies and to decompose the artifact-contaminated data into additional brain signal components such that artifact and brain signal waveforms are maximally independent under the spatial constraint. Thus, in contrast to any other approach to artifact correction SCICA uses both available spatial information about the artifacts and the temporal assumption that artifact and signal waveforms are maximally independent. An iterative algorithm has been introduced in this thesis performing the decomposition by direct evaluation of an ICA contrast function. To make the decomposition computationally effective the dimension of the decomposed data segment, i.e. the number of estimated SCICA components, is always reduced to the minimum necessary. This is also a standard preprocessing step in ICA decomposition. The SCICA approach is not only useful for artifact correction in spontaneous EEG/MEG but seems to be applicable to continuous event-related recordings too. This issue has only been addressed theoretically in the current thesis, though. As an iterative decomposition, SCICA is not fast enough to be employed directly during clinical review of spontaneous EEG/MEG. However, the approach is sufficiently automated to be applied to the whole recording before review or in combination with automated, less time-critical processing such as averaging or mean FFT analysis.

Below, the preselection approach and SCICA are compared with other approaches to artifact correction, i.e. EOG subtraction, projection method, MSEC and ICA. In contrast to EOG subtraction and projection method, both new approaches avoid distortion of brain activity by modeling not only artifact but also brain activity. Contrary to EOG subtraction, neither the preselection approach nor SCICA depend on an external reference signal. The need for an external reference signal is a crucial drawback. On the one hand, reference signals are not available for all types of artifact, e.g. for muscle or line noise. On the other hand, available reference signals such as the EOG or EKG are not always suitable. The EOG may be contaminated by brain activity. The EKG incorporates details about the artifact that are not present in the EEG/MEG.

In comparison to MSEC, preselection and SCICA do not depend on the existence of a dipole source model and are, therefore, applicable to continuous recordings. Contrary to surrogate MSEC, preselection and SCICA not only approximate brain activity but try to estimate it precisely. Nevertheless, optimizing MSEC remains indispensable for artifact correction in averaged event-related data. Alternatively, SCICA or ICA (Jung et al., 2000b) may be applied to continuous event-related recordings before averaging.

In contrast to ICA, no a posteriori visual identification of artifact components is necessary since artifact topographies are known in advance using preselection or SCICA. Visual identification in each epoch of the recording can be very time-consuming especially if artifacts are decomposed into different independent components. As a possible alternative to SCICA, an automated identification of ICA artifact components by topography has been considered in this thesis. Although often successful, the approach fails if eye movement artifacts are decomposed into (multiple) ICA components exhibiting only a low correlation with the predefined artifact subspace. Thus, incorporating known artifact topographies into the estimate of the brain signal subspace seems to be preferable to an a posteriori identification of artifact components by topography.

Determining artifact topographies in advance poses usually no problem. The topography of horizontal and vertical eye movements, blinks and cardiac activity can easily be derived from single or averaged artifact prototypes. Oblique eye movements are expressed as linear combination of horizontal and vertical eye movements. If eye movements or cardiac artifacts are not completely eliminated by artifact correction, the respective artifact may have to be modeled by more than one eigenvector. For muscle artifacts and line noise, on the contrary, it is difficult to find a suitable prototypical pattern. Their topographies may be drawn from an ICA decomposition or the signal subspace of SCICA by visual inspection of the components. Once the topographies are determined, they can be used automatically for artifact correction in any further epoch of the recording. Using one artifact topography for the whole recording implies that the spatial distribution of artifacts is constant. Although this can often be assumed, the spatial distribution of artifacts may vary. Cardiac contamination, for example, is altered if the position of the head relative to the heart changes. The spatial distribution of artifacts in MEG recordings varies if the position of the head relative to the MEG sensors does not remain stable. In SCICA small changes of the spatial distribution of artifacts are compensated by slight variations of the predefined artifact topographies.

ICA and SCICA require that artifact and brain activity are temporally independent. The spatial distribution of the components, however, does not influence the decompositions. The contrary holds for the preselection approach. Temporal independence is in general guaranteed when using sufficiently large data segments. As SCICA incorporates artifact topographies, the requirement of independence is even less restrictive than in ICA. For SCICA to separate simultaneous artifact and brain activity it is sufficient if periods of isolated brain activity exist. ICA needs at least one further artifact in a different context to express its functional independence from the brain activity. Neither ICA (Jung et al., 2000a,b) nor SCICA are suitable, however, if artifact and cerebral waveforms closely resemble and occur (nearly) synchronously only as, for instance, in averaged evoked recordings. In that case SCICA will probably yield a brain signal component whose topography combines the artifact and cerebral spatial distribution and whose waveform represents the synchronous time course of both sources. The waveform corresponding to the predefined artifact topography depicts only the small residual periods of solo artifact activity. SCICA can separate such synchronous sources, however, if their topographies are known in advance.

Both approaches introduced in this thesis present an important improvement over earlier approaches to artifact correction. Above all they are capable of fulfilling the main goal of artifact correction that is to remove artifacts completely without relevant distortion of brain activity. We conclude that SCICA is superior to the preselection approach. Although SCICA is not as fast as preselection it provides a more comprehensive model of brain activity as the brain signal subspace is estimated from the whole data segment. SCICA is easier to use than preselection as no crucial parameters have to be set. Moreover, SCICA seems to be applicable to artifact correction in continuous event-related EEG/MEG recordings too. Both novel approaches may also be utilized in other contexts. One possible application is to trace further occurrences of a particular signal of interest, e.g. an epileptiform spike, by generating a waveform depicting only the signal of interest comparable to the reconstructed artifact waveforms.

9. References

Amari S, Cichocki A, Yang H (1996). A new learning algorithm for blind signal separation. In: Touretzky D, Mozer M, Hasselmo M (eds). Advances in neural information processing systems 8. Cambridge, MA: MIT Press, P757-763.

Amari SI (1998). Natural gradient works efficiently in learning. Neural Computation 10: 251-276.

Antervo A, Hari R, Katila T, Ryhänen T, Seppänen M (1985). Magnetic fields produced by eye blinking. Electroencephalogr. Clin. Neurophysiol. 61: 247-253.

Antoni H (1993). Funktionen des Herzens. In: Schmidt RF, Thews G (eds). Physiologie des Menschen, 25th ed. Berlin: Springer, P461-504.

Barlow JS and Rémond A (1981). Eye movement artifact nulling in EEGs by multichannel on-line EOG subtraction. Electroencephalogr. Clin. Neurophysiol. 52: 418-423.

Barry W and Melvill Jones G (1965). Influence of eye lid movement upon electro-oculographic recording of vertical eye movements. Aerospace Med. 36, 855-858.

Bell AJ and Sejnowski TJ (1995). An information-maximization approach to blind separation and blind deconvolution. Neural Computation 7, 1129-1159.

Berg P (1986). The residual after correcting event-related potentials for blink artifacts. Psychophysiology 23 (3): 354-364.

Berg P and Scherg M (1991a). Dipole models of eye movements and blinks. Electroencephalogr. Clin. Neurophysiol. 79: 36-44.

Berg P and Scherg M (1991b). Dipole modelling of eye activity and its application to the removal of eye artefacts from the EEG and MEG. Clin. Phys. Physiol. Meas. 12, Suppl. A: 49-54.

Berg P and Scherg M (1994). A multiple source approach to the correction of eye artifacts. Electroencephalogr. Clin. Neurophysiol. 90: 229-241.

Brunia CHM, Möcks J, van den Berg-Lenssen MMC, Coelho M, Coles MGH, Elbert T, Gasser T, Gratton G, Ifeachor EC, Jervis BW, Lutzenberger W, Sroka L, van Blokland-Vogelesang AW, van Driel G, Woestenburg JC, Berg P, McCallum WC, Pham DT, Pocock PV, Roth WT (1989). Correcting ocular artifacts in the EEG: a comparison of several methods. J. of Psychophysiology 3: 1-50.

Cardoso JF and Souloumiac A (1993). Blind beamforming for non Gaussian signals. IEE Proceedings-F 140 (6): 362-370.

Cardoso JF and Laheld BH (1996). Equivariant adaptive source separation. IEEE Trans. on Sig. Proc. 44 (12): 3017-3030.

Cardoso JF (1997). Infomax and maximum likelihood for source separation. IEEE Letters on Sig. Proc. 4: 112-114.

Cardoso JF (1998). Blind signal separation: statistical principles. Proc. of the IEEE 9 (10): 2009-2025.

Comon P (1992). Independent component analysis. In: Lacoume JL (ed). Higher-order statistics. Amsterdam: Elsevier, P29-38.

Comon P (1994). Independent component analysis, a new concept? Signal Process. 36: 287-314.

Cooper R, Osselton JW, Shaw JC (1984). Elektroenzephalographie, 3rd ed. Stuttgart: Gustav Fischer Verlag.

Corby JC and Kopell BS (1972). Differential contributions of blinks and vertical eye movements as artifacts in EEG recording. Psychophysiology 9 (6): 640-644.

Croft RJ and Barry RJ (2000). EOG correction of blinks with saccade coefficients: a test and revision of the aligned-artifact average solution. Clin. Neurophysiol. 3: 444-455.

Dongarra JJ, Bunch JR, Moler CB, Stewart GW (1979). LINPACK users' guide. Philadelphia: SIAM.

Ebersole JS (1991). EEG dipole modelling in complex partial epilepsy. Brain Topography 4: 113-123.

Ebner A, Sciarretta G, Epstein CM, Nuwer M (1999). EEG instrumentation. In: Deuschl G, Eisen A (eds). Recommendations for the practice of clinical neurophysiology: guidelines of the international federation of clinical neurophysiology, 2nd ed. Electroencephalogr. Clin. Neurophysiol., Suppl. 52. Elsevier, P7-10.

Elbert T, Lutzenberger W, Rockstroh B, Birbaumer N (1985). Removal of ocular artifacts from the EEG – a biophysical approach to the EOG. Electroencephalogr. Clin. Neurophysiol. 60: 455-463.

Fortgens C and de Bruin MP (1983). Removal of eye movement and ECG artifacts from the non-cephalic reference EEG. Electroencephalogr. Clin. Neurophysiol. 56: 90-96.

Gasser T, Sroka L, Möcks J (1985). The transfer of EOG activity into the EEG for eyes open and closed. Electroencephalogr. Clin. Neurophysiol. 61: 181-193.

Gasser T, Sroka L, Möcks J (1986). The correction of EOG artifacts by frequency dependent and frequency independent methods. Psychophysiology 23 (6): 704-712.

Girolami M (1998). An alternative perspective on adaptive independent component analysis algorithms. Neural Computation 10 (8): 2103-2114.

Girton DG and Kamiya J (1973). A simple on-line technique for removing eye movement artifacts from the EEG. Electroencephalogr. Clin. Neurophysiol. 34: 212-216.

Golub GH and van Loan CF (1996). Matrix computations, 3rd ed. Baltimore, London: Johns Hopkins University Press.

Gratton G, Coles MGH, Donchin E (1983). A new method for off-line removal of ocular artifact. Electroencephalogr. Clin. Neurophysiol. 55: 468-484.

Hari R (1993). Magnetoencephalography as a tool of clinical neurophysiology. In: Niedermayer E, Lopes da Silva F (eds). Electroencephalography. Basic principles, clinical applications, and related fields. Baltimore: Williams & Wilkins, P1035-1061.

Hector ML (1980). EEG recording. London: Butterworths.

Hillyard SA and Galambos R (1970). Eye movement artifact in the CNV. Electroencephalogr. Clin. Neurophysiol. 28: 173-182.

Huotilainen M, Ilmoniemi RJ, Tiitinen H, Lavikainen J, Alho K, Kajola M, Näätänen R (1995). The projection method in removing eye-blink artefacts from multichannel MEG measurements. In: Baumgartner C, Deecke L, Stroink G, Williamson SJ (eds). Biomagnetism: fundamental research and clinical applications. Amsterdam: Elsevier Science, IOS Press, P363-367.

Hyvärinen A (1999). Survey on independent component analysis. http://www.cis.hut.fi/~aapo/

Hyvärinen A and Oja E (2000). Independent component analysis: algorithms and applications. Neural Networks 13: 411-430.

Iacono WG and O'Toole DM (1987). An evaluation of different techniques for removing eye-blink artifact from visual evoked response recordings. Psychophysiology 24 (4): 487-497.

Ille N, Berg P, Scherg M (1997). A spatial components method for continuous artifact correction in EEG and MEG. Biomed. Tech. 42, Suppl. 1: 80-83.

Ille N, Beucker R, Scherg M (2001). Spatially constrained independent component analysis for artifact correction in EEG and MEG. Neuroimage 13: S159.

Ilmoniemi RJ (1992). Method and apparatus for separating the different components of evoked response and spontaneous activity brain signals as well as of signals measured from the heart. Finnish patent application number 925461.

Jervis BW, Ifeachor EC, Allen EM (1988). The removal of ocular artefacts from the electroencephalogram: a review. Med. Biol. Eng. Comput. 26: 2-12.

Jousmäki V and Hari R (1996). Cardiac artifacts in magnetoencephalogram. J. Clin. Neurophysiol. 13 (2): 172-176.

Jung TP, Humphries C, Lee TW, Makeig S, McKeown MJ, Iragui V, Sejnowski TJ (1998). Extended ICA removes artifacts from electroencephalographic recordings. In: Jordan MI, Kearns MJ, Solla SA (eds). Advances in neural information processing systems 10. Cambridge MA: MIT Press, P894-900.

Jung TP, Makeig S, Humphries C, Lee TW, McKeown MJ, Iragui V, Sejnowski TJ (2000a). Removing electroencephalographic artifacts by blind source separation. Psychophysiology 37: 163-178.

Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ (2000b). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clin. Neurophysiol. 111: 1745-1758.

Katila T, Maniewski R, Poutanen T, Varpula T, Karp PJ (1981). Magnetic fields produced by the human eye. J. Appl. Phys. 52 (3): 2565-2571.

Kendall M and Stuart A (1977). The advanced theory of statistics, vol.1, distribution theory, 4th ed. London: Charles Griffin & Company.

Lee TW, Girolami M, Sejnowski TJ (1999). Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources. Neural Computation 11 (2): 409-433.

Lins OG, Picton TW, Berg P, Scherg M (1993). Ocular artifacts in recording EEGs and event-related potentials II: source dipoles and source components. Brain Topography 6 (1): 65-78.

Makeig S, Bell AJ, Jung TP, Sejnowski TJ (1996). Independent component analysis of electroencephalographic data. In: Touretzky D, Mozer M, Hasselmo M (eds). Advances in neural information processing systems 8. Cambridge MA: MIT Press, P145-151.

Matsuo F, Peters JF, Reilly EL (1975). Electrical phenomena associated with movements of the eyelid. Electroencephalogr. Clin. Neurophysiol. 38: 507-511.

Meyer CD (2000). Matrix analysis and applied linear algebra. Philadelphia PA: SIAM.

Mosher JC, Lewis PS, Leahy R (1992). Multiple dipole modeling and localization from spatio-temporal MEG data. IEEE Trans. Biomed. Eng. 39: 541-557.

Nadal JP and Parga N (1994). Non-linear neurons in the low noise limit: a factorial code maximizes information transfer. Network 5: 565-581.

Nelder JA and Mead R (1965). A simplex method for function minimization. Computer J. 7: 308-313.

Nuwer MR, Lehmann D, Lopes da Silva F, Matsuoka S, Sutherling W, Vibert JF (1999). IFCN guidelines for topographic and frequency analysis of EEGs and EPs. In: Deuschl G, Eisen A (eds). Recommendations for the practice of clinical neurophysiology: guidelines of the international federation of clinical neurophysiology, 2nd ed. Electroencephalogr. Clin. Neurophysiol., Suppl. 52. Elsevier, P15-20.

Overton DA and Shagass C (1969). Distribution of eye movement and eyeblink potentials over the scalp. Electroencephalogr. Clin. Neurophysiol. 27: 546.

Papoulis A (1991). Probability, random variables, and stochastic processes, 3rd ed. New York: McGraw-Hill.

Pham DT, Garat P, Jutten C (1992). Separation of a mixture of independent sources through a maximum likelihood approach. In: Vandewalle J, Boite R, Moonen M, Oosterlinck A (eds). Signal processing VI: theories and applications. Elsevier Science, P771-774.

Picton TW, van Roon P, Armilio ML, Berg P, Ille N, Scherg M (2000a). The correction of ocular artifacts: a topographic perspective. Clin. Neurophysiol. 111: 53-65.

Picton TW, van Roon P, Armilio ML, Berg P, Ille N, Scherg M (2000b). Blinks, saccades, extraocular muscles and visual evoked potentials (reply to Verleger). J. of Psychophysiology 14 (4): 210-217.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992). Numerical recipes in C, 2nd ed. Cambridge: Cambridge University Press.

Scherg M and von Cramon D (1985). Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. Electroencephalogr. Clin. Neurophysiol. 62: 32-44.

Scherg M (1990). Fundamentals of dipole source analysis. In: Grandori F, Hoke M, Romani GL (eds). Auditory evoked magnetic fields and electric potentials. Advances in audiology 6. Basel: Karger, P40-69.

Scherg M, Linden DEJ, Muckli L, Roth R, Drüen K, Ille N, Linek S, Zanella FE, Singer W, Göbel R (1999). Combining MEG with fMRI in studies of the human visual system. In: Yoshimoto T, Kotani M, Kuriki S, Karibe H, Nakasato N (eds). Recent advances in biomagnetism. Sendai: Tohoku University Press, P129-132.

Schwind J and Dormann WU (1986). Off-line removal of ocular artifacts from event-related potentials using a multiple linear regression model. Int. J. Psychophysiol. 4: 203-208.

Semlitsch HV, Anderer P, Schuster P, Presslich O (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. Psychophysiology 23 (6): 695-703.

Van den Berg-Lenssen MMC, Brunia CHM, Blom JA (1989). Correction of ocular artifacts in EEGs using an autoregressive model to describe the EEG; a pilot study. Electroencephalogr. Clin. Neurophysiol. 73: 72-83.

Verleger R, Gasser T, Möcks J (1982). Correction of EOG artifacts in event-related potentials of the EEG: aspects of reliability and validity. Psychophysiology 19 (4): 472-480.

Vigário RN (1997). Extraction of ocular artefacts from EEG using independent component analysis. Electroencephalogr. Clin. Neurophysiol. 103: 395-404.

Vigário RN, Jousmäki V, Hämäläinen M, Hari R, Oja E (1998). Independent component analysis for identification of artifacts in magnetoencephalographic recordings. In: Jordan MI, Kearns MJ, Solla SA (eds). Advances in neural information processing systems 10. Cambridge MA: MIT Press, P229-235.

Whitton JL, Lue F, Moldofsky H (1978). A spectral method for removing eye movement artifacts from the EEG. Electroencephalogr. Clin. Neurophysiol. 44: 735-741.

Woestenburg JC, Verbaten MN, Slangen JL (1983). The removal of the eye-movement artifact from the EEG by regression analysis in the frequency domain. Biol. Psychol. 16: 127-147.

10. Abbreviations

EEG	electroencephalography, electroencephalogram, electroencephalographic
EKG	electrocardiogram
EOG	electro-oculogram
HEOG	horizontal electro-oculogram
ICA	independent component analysis
MEG	magnetoencephalography, magnetoencephalogram, magnetoencephalographic
MSEC	multiple source eye correction
PCA	principal component analysis
pdf	probability density function
SCICA	spatially constrained independent component analysis
SNR	signal-to-noise ratio
SVD	singular value decomposition
VEOG	vertical electro-oculogram

11. Mathematical notation and symbols

a	small letters in italics denote variables or scalars
$\mathbf{a}$	small boldface letters denote column vectors
$\mathbf{A}$	capital boldface letters denote matrices
$\mathbf{a} \in \Re^{m}$	vector of m real elements
$\mathbf{A} \in \Re^{m \times n}$	matrix of m rows and n columns; each matrix element is a real number
$\mathbf{A}(:,1:n)$	matrix composed of all rows and columns 1 to n of matrix $\mathbf{A}$
$\mathbf{A}^{\mathrm{T}}$, $\mathbf{a}^{\mathrm{T}}$	transpose of matrix $\mathbf{A}$ or column vector $\mathbf{a}$
$\mathbf{A}^{-1}$	inverse of a nonsingular square matrix $\mathbf{A}$
$\mathbf{A}^{-\frac{1}{2}}$	square root of the inverse of $\mathbf{A}$; $\mathbf{A}^{-\frac{1}{2}}\mathbf{A}^{-\frac{1}{2}} = \mathbf{A}^{-1}$
$\mathbf{A}^{\dagger}$	Moore-Penrose pseudo-inverse of matrix $\mathbf{A}$
$\Vert \mathbf{a} \Vert$	length of vector $\mathbf{a}$, $\Vert \mathbf{a} \Vert = \sqrt{\mathbf{a}^{T}\mathbf{a}}$
$\Vert \mathbf{A} \Vert_{2}$	2-norm of matrix $\mathbf{A}$
$\mathrm{Cum}[a_1, ..., a_r]$	r^{th} order cumulant
$\det(\mathbf{A})$	determinant of square matrix $\mathbf{A}$
$\mathrm{diag}(\mathbf{a})$	square matrix: diagonal is equal to $\mathbf{a}$, off-diagonal elements are zero
$\mathrm{diag}(a_1, ..., a_n)$	square matrix: diagonal is equal to $a_1, ..., a_n$, off-diagonal elements are zero
$\mathrm{diag}(\mathbf{A})$	vector built from the diagonal of matrix $\mathbf{A}$
$\mathrm{E}\{\mathbf{a}\}$	expectation, i.e. average of elements of $\mathbf{a}$
$\mathbf{e}_i$	i^{th} canonical vector having 1 at the i^{th} position and zeros elsewhere
$\mathrm{g}'(x)$	first derivative of function $\mathrm{g}(x)$
$\mathbf{I}$	identity matrix, i.e. $\mathrm{diag}(1,\ldots,1)$
$\ln(x)$	natural logarithm of x
$\mathrm{range}(\mathbf{A})$	vector space spanned by the column vectors of matrix $\mathbf{A}$
$\mathrm{rank}(\mathbf{A})$	dimension of vector space spanned by the column vectors of matrix $\mathbf{A}$
$\mathrm{sign}(a)$	signum function, 1 for $a > 0$, -1 for $a < 0$
$\mathrm{span}(\mathbf{a}_1, ..., \mathbf{a}_n)$	vector space spanned by the column vectors $\mathbf{a}_1, ..., \mathbf{a}_n$
ϕ	contrast function
ϕ^{o}	orthogonal contrast function
$\propto$	proportional to

12. Acknowledgements

First, I want to thank my supervisor Prof. Dr. Michael Scherg, University of Heidelberg, for directing me to the interesting topic of artifact correction and for providing me with the necessary background of spatial filtering. I appreciate his valuable comments on this manuscript and my publications.

Moreover, I am very grateful to Prof. Dr. Reinhard Männer and assistant professor Dr. Jürgen Hesser for readily accepting and supporting my thesis at an advanced stage. In addition to reviewing this manuscript and a not yet published paper about SCICA, Jürgen provided any support needed and contributed with important questions and discussions.

I am very indebted to Dr. Patrick Berg, University of Konstanz, and Dr. med. Alexander Gutschalk, University of Heidelberg, for thoroughly proof-reading my manuscript and publications. Both showed great interest in this work and provided helpful comments and interesting examples for artifact correction.

Another person important to my research was Dr. Roland Beucker. Roland was always willing to discuss my ideas on SCICA and critically reflected upon the mathematical notation in this dissertation. Moreover, I am grateful to him for his fruitful comments on the upcoming SCICA paper.

The EEG examples presented in this thesis are partially by the courtesy of Prof. Dr. John Ebersole, University of Chicago. Further examples have been derived from the numerous EEGs and MEGs recorded at the Department of Neurology, University of Heidelberg. I particularly thank the technical assistants at the Department of Neurology: Barbara Burghardt, Uwe Gollner-Nohlen, Martina Kirsch, Annette Opgenorth, Esther Tauberschmidt, Claudia Thor, and Theda Unger who made the recordings.

I am also grateful to Prof. Dr. Terence Picton, University of Toronto, who allowed me to contribute to two of his papers.

Moreover, I thank all other fellow researchers and co-workers at the Division of Biomagnetism, Department of Neurology, University of Heidelberg, and my colleagues at MEGIS Software GmbH.

Finally, I am extremely grateful to my parents who gave me enormous support at all times.

This research was supported in part by a LGFG grant of the University of Heidelberg.