Technische Universität Berlin /
Physikalisch-Technische Bundesanstalt

Florian Kohl

Blind separation of dependent source signals for MEG sensory stimulation experiments

Blind separation of dependent source signals for MEG sensory stimulation experiments

vorgelegt von
Dipl.-Ing.

Florian Kohl

von der Fakultät IV, Elektrotechnik und Informatik
der Technischen Universität Berlin

zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften
- Dr.-Ing. -
genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr. Klaus-Robert Müller, TU Berlin
Berichter:	Prof. Dr. Reinhold Orglmeister, TU Berlin
Berichter:	Prof. Dr. Jens Haueisen, TU Ilmenau

Tag der wissenschaftlichen Aussprache: 26. April 2013

Berlin 2013
D83

Abstract

Sensory stimulus evoked magnetoencephalography (MEG) data are known to be highly impacted by background brain and technical noise sources. In order to enhance the signal quality, sophisticated signal processing is necessary and becomes crucial when looking at single-trials. Often, linear decomposition techniques are utilized. Among many methods, those under the framework independent component analysis (ICA) are most popular. Linear decomposition aims at finding an unmixing matrix such that noise and signal sources can be recovered from the recorded data. ICA is used to reduce artifacts or, optimally, to obtain a clear copy of the evoked MEG source signal of interest.

Specifically, in a sensory stimulation experimental setting, the underlying source signal of interest occurs after presenting a sensory stimulus. An ICA decomposition of evoked data often yields a component that shows this behavior; it is tempting to refer to this ICA component as *the* evoked source signal recovered from the recordings. However, ICA assumes independence among the underlying source signals. If more than one source signal is evoked by one or more stimuli, the ICA independence assumption can be violated.

In the context of evoked dependent sources, this thesis investigates ICA and proposes a novel blind separation method. As starting point, a virtual evoked MEG experiment with adjustable source signal dependencies is designed and used for assessment of various decomposition methods. Furthermore, an audio-visual MEG data experiment is designed for a real world test. Rather surprisingly, it is demonstrated that ICA is able to recover highly dependent source signals - for specific source signal settings. Along this line, the use of ICA for decomposition as well as for subspace identification is discussed. Subsequently, the novel time domain shifted factor analysis (TDSFA) technique is proposed. In particular, TDSFA is based on a Taylor series expansion of the shifted factor analysis (SFA) model, which is unique without assuming independence among the source signals. The relation to a classic unique trilinear technique is investigated and it is shown that the trilinear technique can be very sensitive to shifts. Indeed, TDSFA accounts for shifts and is shown to be more suited for decomposition of evoked MEG data.

Utilizing actual as well as virtual MEG data, the results show that ICA and other state of the art techniques can fail. The results suggest that the novel TDSFA technique has high potential as a decomposition technique in the context of evoked MEG source signals.

Zusammenfassung

Sensorisch evozierte Magnetenenzephalographie-Daten (MEG-Daten) sind stark durch das Hintergrundrauschen im Gehirn und durch technische Rauschquellen gestört. Um die Signalqualität zu verbessern sind moderne Verfahren der Signalverarbeitung notwendig, insbesondere für eine Single-Trial-Analyse.
Für die Verbesserung der MEG-Signalqualität werden oft lineare Zerlegungsmethoden eingesetzt, wobei die Methoden der Independent Component Analysis (ICA) eine weite Verbreitung gefunden haben.
Lineare Zerlegungsmethoden versuchen anhand gemischter Daten eine Entmischungsmatrix zu gewinnen, um durch eine lineare Entmischung die Quellsignale wiederherzustellen. ICA wird für MEG-Daten auch benutzt, um die Anzahl der Rauschquellen zu reduzieren - im optimalen Fall, um ein gewünschtes MEG-Quellsignal unverrauscht zu rekonstruieren.
Für den Fall der sensorisch evozierten MEG-Daten treten die gewünschten MEG-Quellsignale erst nach erfolgter sensorischer Reizung auf. Eine ICA-Zerlegung evozierter MEG-Daten zeigt oft eine Komponente mit entsprechendem zeitlichen Verhalten; man ist versucht diese Komponente als *das originale* evozierte MEG-Quellsignal zu interpretieren. ICA nimmt jedoch die statistische Unabhängigkeit der Quellsignale an. Werden durch eine sensorische Stimulation mehrere Quellsignale im Gehirn evoziert, so kann die Unabhängigkeitsannahme verletzt sein.
In dieser Arbeit wird in dem Zusammenhang abhängiger evozierter MEG-Quellsignale ICA untersucht und eine neue Zerlegungsmethoden vorgestellt. Es wird eine virtuelle Simulationsumgebung mit steuerbaren Quellsignalabhängigkeiten vorgestellt und zur Untersuchung der verschiedenen Zerlegungsverfahren eingesetzt. Er wird gezeigt, dass ICA in Sonderfällen stark abhängige Signale trennen kann. Anhand dieser Erkenntnis wird der Nutzen von ICA zur Signaltrennung und zur Unterraumanalyse von evozierten MEG-Daten diskutiert. Ein reales MEG-Experiment zeigt das Verhalten der Methoden auch für den Fall von echten evozierten MEG-Daten.
Als neue Methode wird Time Domain Shifted Factor Analysis (TDSFA) vorgestellt. TDSFA basiert auf einer Taylorreihenentwicklung des Modells der Shifted Factor Analysis (SFA), welches eindeutig identifizierbar ist ohne die Unabhängigkeit der Quellsignale annehmen zu müssen. Die Verbindung zu den etablierten trilinearen Methoden wird untersucht, und es wird gezeigt, dass die etablierten trilinearen Methoden Schwierigkeiten mit zeitlicher Signalverschiebung haben. TDSFA bezieht Zeitverschiebung in das Modell ein und ist für die lineare Zerlegung von evozierten MEG-Daten besser geeignet.
Für echte und simulierte evozierte MEG-Daten zeigen die Ergebnisse, dass ICA und andere moderne Methoden falsche Ergebnisse liefern können. Die neue Zerlegungsmehtode TDSFA liefert bessere Ergebnisse und weist hohes Potential im Kontext evozierter MEG-Daten auf.

Acknowledgment

This work is based on a fruitful cooperation between the Physikalische-Technische Bundesanstalt (PTB) and the Electronics and Medical Signal Processing (EMSP) Group at the TU Berlin. Financial support from the Deutsche Forschungsgemeinschaft DFG (Grant OR 99/4-1 and Grant BA 1222/6-1) is gratefully acknowledged.

I would like to thank the many persons who have helped me along the way in researching and writing this thesis.
Among them, I wish to express my sincere gratitude to Prof. Dr. Orglmeister, who supervised me at the TU Berlin. Thank you for the opportunity of working in this vivid field of research and for all the time you offered to advance our project. I highly acknowledge the input of Prof. Dr. Kolossa; all her effort and her valuable thoughts were essential.
I owe debt of gratitude to the support and inspiration of many scientists from the PTB. Namely, Dr. Elster, Dr. Wübbeler and Dr. Bär positively influenced this thesis through many enriching discussions. Thank you for sharing your experience throughout these years.
I warmly thank Dr. Sander-Thömmes and Dr. Trahms for their competent support in recording MEG data.
My special thanks goes to all of my students and my colleagues worldwide, who I supervised and who supervised me. Lastly, I would like to thank my parents, my brother and my family, in particular Johanna, for the endless support.

To all of you, all the best and good luck on the way of improving signal processing for the sake of humanity.

Florian Kohl

Contents

Abbreviations

AEF	Auditory Evoked Field
ALS	Alternating Least Squares
CANDECOMP	Canonical Decomposition
CP	CANDECOMP/PARAFAC
DTD	Direct Trilinear Decomposition
ECD	Equivalent Current Dipole
EEG	Electroencephalography
EPSP	Excitatory Postsynapic Potential
ERD	Event Related Desynchronization
ERF	Event Related Field
ERP	Event Related Potential
ERS	Event Related Synchronization
fMRI	Functional Magnetic Resonance Imaging
ICA	Independent Component Analysis
IPSP	Inhibitory Postsynapic Potential
LLS	Linear Least Squares
MEG	Magnetoencephalography
MI	Mutual Information
N100	ERF's prominent deflection at 100 ms
NIRS	Near Infrared Spectroscopy
PARAFAC	Parallel Factor Analysis
PCA	Principal Component Analysis
PET	Positron Emission Tomography
SEF	Somatosensory Evoked Field
SFA	Shifted Factor Analysis
SQUID	Super Conducting Quantum Interference Device
TLD	Trilinear Decomposition
VEF	Visual Evoked Field

MEG data	Recordings of an MEG experiment, a mixture of MEG sources
Actual MEG data	Data from a real world experiment
Virtual MEG data	Data from a computer simulation experiment
Source	Random variable that contributes to (mixed) data
Source signal	Source with temporal structure

Notations

$\mathbf{A}$	Mixing matrix
$\mathbf{W}$	Unmixing matrix
$\mathbf{V}$	Whitening matrix
$\mathbf{O}$	Whitening mixing matrix product
$\mathrm{E}[\cdot]$	Expectation
$\mathrm{kurt}(\cdot)$	Kurtosis
λ	Eigenvalue
$\mathrm{std}(\cdot)$	Standard deviation
$\mathrm{var}(\cdot)$	Variance
$\mathbf{s}$	Source signal
$\mathbf{x}$	Mixed signal
$\mathbf{z}$	Whitened signal
$\mathbf{u}$	Estimated signal
x	Random variable function
x_0	Instance of random variable's range
$P(x)$	Probability mass function
$p(x)$	Probability density function
$\underline{\mathbf{N}}$	Third order tensor
N_{ijk}	Element of a third order tensor
$\mathbf{N}_{(i)}$	Matrix from ith tensor unfolding
$\mathbf{N}_i$	Ith column vector of a matrix
$\lvert \cdot \rvert$	Absolute value
$\lvert\lvert \cdot \rvert\rvert$	Euclidean norm
$\lvert\lvert \cdot \rvert\rvert_F$	Frobenius norm
$\circ$	Hadamard product
$\otimes$	Kronecker product
$\odot$	Khatri-Rao product
$(\cdot)^{-1}$	Matrix inverse
$(\cdot)^{\#}$	Matrix pseudoinverse
$(\cdot)^{T}$	Matrix transpose
$\mathrm{diag}(\cdot)$	Vector to diagonal matrix operator
$\det(\cdot)$	Matrix determinant
$\mathbf{1}$	Full matrix with ones
$\mathbf{I}$	Identity matrix
$\mathbf{P}$	Permutation matrix
$\mathbf{D}$	Diagonal matrix

$\mathbf{Q}$	Orthogonal matrix
$\mathbf{E}$	Eigenvector matrix
$\mathbf{\Lambda}$	Eigenvalue matrix

1 Introduction

The functioning of the human brain is of great general interest. It has an impact on many disciplines such as medicine, engineering and psychology [Zsc02, Mün00]. Brain signals are used for medical diagnostics, to quantify psychological effects or as input for brain computer interfaces [Nun81, Mül08]. A key technique for understanding brain functioning is the analysis of brain signals related to sensory stimulations. In a sequence of trials, one or multiple stimuli are presented to a human subject and the evoked brain signals are recorded by a measuring device [Pic95, Mün00].
Due to its excellent temporal resolution, magnetoencephalography (MEG) is next to electroencephalography (EEG)[1] often used to record brain signals. MEG is a non-invasive technique that records magnetic fields at many positions distributed in space [Häm93]. The recorded fields can be related to the evoked electrophysiological brain signals. However, many sources of interference contribute to the MEG data. Although many methods have been proposed [Vig09, Pic95], the evoked signal recovery task remains highly challenging. Specifically, ongoing brain activities and external noise sources can be by magnitudes larger than evoked activity. Furthermore, spatial, temporal and frequency domains of noise interferers and evoked signals often overlap; the evoked activity may not be simply recovered by standard filtering techniques.
Due to the low signal-to-noise ratio, evoked MEG data are often averaged over trials [Vig09]. The result is referred to as the event related field (ERF). However, this technique assumes that evoked activity has a fixed latency between onset time of stimulation and evoked signal for all trials. It neglects that evoked signals can vary from trial to trial [Tru02]. Interesting and potentially relevant single-trial variations may not be recovered [Sal04, Jun01]. More recently, single-trial analysis emerged, aiming at non-averaged ERFs that follow each stimulus [Tan05, Sal04, Jun01]. A promising technique is to decompose the recorded data prior to single-trial analysis for the sake of artifact reduction or, optimally, to obtain a clear copy of the evoked signal [Tan05]. The idea behind linear decomposition techniques is that the recorded MEG data consist of linearly superimposed contributions from many underlying source signals at each recording channel. Among the sources, some are of interest, many more are merely background brain and technical noise sources.
To date, independent component analysis (ICA) is the most prominent tool for MEG data decomposition. ICA is a statistical technique that aims to linearly decompose multivariate signals into underlying independent source signals. ICA has been very successful in removing noise sources, such as those introduced by muscle contractions, eye movement, electrical line or heartbeat. These noise sources are often statistically independent. In contrast, if more than one signal is evoked by one or more stimuli, this assumption may no longer hold for evoked brain signals.

[1] In this work, all considerations are limited to MEG data; results directly apply to EEG data.

1.1 Scope

This thesis addresses the recovery of evoked brain signals by means of mathematical decomposition techniques. It is tempting to use ICA to separate the evoked signals from MEG data. Often, the results look reasonable, i.e. often, dipolar magnetic field patterns suggest a successful separation. However, evoked signals may be dependent and the key assumption of ICA may be violated. Hence, the separation of dependent evoked signals shall be in focus. Specifically, the hypothesis that evoked signals can be dependent is tested and ICA's robustness against the violation of independence is investigated. To what extent evoked signals are *statistically* dependent and to what extent ICA can still be used is currently not well known. It is intrinsic to MEG data that mixing introduces dependency, even in the case of independent source signals. Thus, the use of ICA for the recovery of single-trial evoked signals may be still of use and needs to be assessed by virtual and actual experiments.
Subsequently, alternative decomposition techniques next to ICA are discussed. It is intended to design an algorithm without resorting to different contrasts such as sparsity, non-negativity, non-frequency overlap or others. Contrarily, intrinsic features of evoked MEG data are investigated, that allow a *unique* decomposition. By employing a multilinear model, the inherent structure of evoked experiments is used, yielding uniqueness, without resorting to objectives that may be valid only for a subset of signals.
In summary, four main questions are investigated:

- Can evoked source signals be *statistically* dependent?
- Can ICA separate evoked dependent source signals?
- Can the subspace of evoked dependent source signals be recovered?
- Can evoked source signals be recovered from this subspace?

1.2 Contributions

This thesis contributes with a new method for the separation of MEG data with underlying evoked dependent MEG source signals. A new virtual MEG experiment and a new actual MEG experiment are introduced that can be used as a basis for evaluating the separation performance of different methods. This is important as the ground truth of actual MEG data is not known in general. It is shown that ICA *can* separate dependent source signals in special cases. Evoked dependent source signals are shown to be not separable by ICA in general. The near dipole effect shows that a dipolar pattern does not ensure the success of ICA.
Furthermore, the identification of the evoked and possibly dependent source signal subspace is discussed. The recovered subspace is shown to be separable by the proposed method. For both actual and simulated MEG data, the separation performance is significantly improved using the novel method in contrast to using ICA.

Publications of this thesis are listed below.

Journal articles

- Kohl, F., Wübbeler, G., Kolossa, D., Bär, M., Orglmeister, R. and Elster, C. 2010. Shifted factor analysis for the separation of evoked dependent MEG signals, *Physics in medicine and biology*, 55:4219–4230.
- Sander, T.H., Knösche, T.R., Schlögl, A., Kohl, F., Wolters, C.H., Haueisen, J. and Trahms, L. 2010. Recent advances in modeling and analysis of bioelectric and biomagnetic sources. *Biomedizinische Technik*, 55:65–76.

Conference articles

- Kohl, F., Wübbeler, G., Sander, T., Trahms, L., Kolossa, D., Orglmeister, R., Elster, C. and Bär, M. 2008. Performance of ICA for dependent sources using synthetic stimulus evoked MEG data. *Proc. DGBMT-Workshop Biosignalverarbeitung*, 32–35.
- Kohl, F., Wübbeler, G., Kolossa, D., Orglmeister, R., Elster, C. and Bär, M. 2008. Performance of ICA for MEG data generated from subspaces with dependent sources. *Proc. ECIFMBE*, 22:1281–1285.
- Kohl, F., Wübbeler, G., Kolossa, D., Elster, C., Bär, M., Orglmeister, R. 2009. Non-independent BSS: a model for evoked MEG signals with controllable dependencies, *Proc. ICA*, lecture notes in computer science, 5441:443–450.
- Kohl, F. and Wübbeler, G. and Kolossa, D. and Elster, C. and Bär, M. and Orglmeister, R. 2010. Noise adjusted PCA for finding the subspace of evoked dependent signals from MEG data. *Proc. LVA*, lecture notes in computer science, 6365:442–449.
- Ghaemi, D., Kohl, F. and Orglmeister, R. 2010. Classifying ICA components of evoked MEG data. *Proc. BMT*, 55:302–305.

1.3 Overview

In Chapter 2, the basis of evoked MEG is introduced. Chapter 3 presents the framework of ICA. In Chapter 4, ICA is investigated in the context of evoked brain signals, using a new virtual as well as a new actual experiment. In Chapter 5, a novel blind separation technique is introduced that is based on multilinear models. Furthermore, Chapter 5 discusses alternative separation techniques and tackles the identification of the evoked signal subspace. Chapter 6 provides a concluding summary.

2 Evoked brain signals

Evoked brain signals give evidence of a direct link between stimulation and neuronal activity. An evoked signal is a response of the brain to an external stimulus, i.e. the flow of ions that result in magnetic fields, which are recorded by MEG sensors.
MEG recordings can provide non-invasive access to evoked signals. These signals are used for various tasks, such as evaluating cognitive paradigms or functional brain mapping. Figure 2.1 depicts the anatomy of the cerebral cortex with brain regions and associated topological naming. For example, auditory stimulations evoke responses in the primary auditory cortex. [Pan88] demonstrated that different acoustic frequencies are mapped to distinct regions therein. Brain signals related to sensory stimulation experiments can further be evoked by visual stimulations, by stimulation of the tactile as well as the olfactory and gustatory senses. The evoked responses are often modulated by attention, emotions, vigilance and cooperation of the subject undergoing testing [Mün00, Hai64].
The exact process of evoked activity is extremely complicated to reconstruct, as the brain consists of more than 10^{10} neurons that are largely interconnected [Häm93]. A virtual copy of the brain, imaging all relevant ion flows and neurons involved during stimulation, is desirable but not feasible to date. An insight into brain processing after stimulation is only incompletely provided and is dependent on the acquisition technique in use.

2.1 Acquisition techniques

Acquisition techniques may be classified into invasive and non-invasive, among which non-invasive recording is clearly preferable. However, it may provide access only to indirect information.
Among the non-invasive methodologies, positron emission tomography (PET) [Ter75], functional magnetic resonance imaging (fMRI) [Oga90], near infrared spectroscopy (NIRS) [Jöb77] as well as electroencephalography (EEG) [Ber29] and magnetoencephalography (MEG) [Coh68] are regularly used.
PET, NIRS and fMRI focus on blood flow and blood oxygenation effects which are linked to neuronal activity [Roy90]. These techniques offer high spatial resolution, however, the exact relation between neuronal activity and vascular activity remains complex [Cae03, Kim03]. For the study of evoked signal dynamics, one of the main concerns is the low temporal resolution of these techniques, which is in the range of seconds.
In contrast, EEG and MEG offer high temporal resolution in the ms range and cover the typical frequency content of evoked signals. Hence, EEG and MEG are preferable to study the dynamics

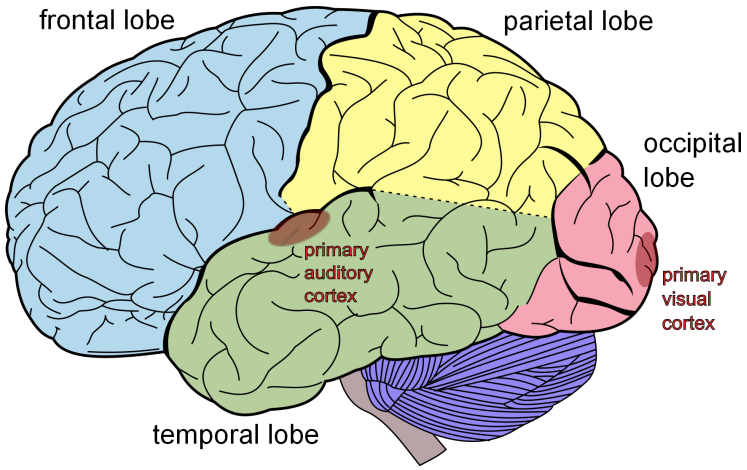


Figure 2.1: Topographical overview of the main cortical areas of the human brain. The primary visual cortex and the primary auditory cortex are highlighted. These areas show prominent activity 100 ms after stimulation; the so-called N100/P100 instant has been studied intensively [Fra03].

of evoked responses [Har05, Häm93]. An EEG records electrical potentials by using multiple electrodes fixed to the scalp. The recorded potentials arise from volume conducting currents that originate from the primary currents in the neurons. An MEG records resulting magnetic fields from primary and secondary currents using multiple magnetic sensors placed near the scalp. In a spherical approximation of the head, only primary currents have to be considered [Sar87].
EEG recordings are affected by inhomogeneities of the head. As a consequence, MEG field maps are often spatially less smeared [Har05]. However, MEG is insensitive to radially orientated currents in the spherical model. Hence, both methods may be regarded as providing complementary information. Indeed, as the neocortex is folded to sulci and gyri both techniques offer their advantages. EEG is most sensitive to the radial currents that are prominent in the gyri and MEG is most sensitive to tangential currents that are prominent in the sulci. Here, the MEG recording modality is investigated; the MEG specific advantages as well as disadvantages have to be kept in mind for making conclusions about the decomposition results. For studying brain functioning, more than one acquisition technique should be considered. Benefits from combining techniques have been recently reported [Dal93, Zav09].

2.1.1 Magnetoencephalography

Magnetic fields that result from evoked neuronal currents in the brain are very weak. They range from a few to several hundreds of femto Tesla. In comparison, the earth magnetic field is about 8 orders larger in magnitude.
As a consequence, in order to record an MEG, external fields have to be attenuated. Commonly, a magnetic shielded room is used, such as the one depicted in Fig. 2.2. Passive shielding often consists of several layers of mu-metal and aluminum. AC coils are used for active shielding

Figure 2.2: In- and outside of a magnetically shielded room for magnetoencephalography recordings. During measurement, doors are closed. Layers of aluminum and mu-metal yield a factor of $10^3 - 10^5$ magnetic field attenuation depending on the spectral content of the noise. The interior of the cabin has to be as magnetically silent as possible. Mirrors bring visual stimulation from a beamer that is placed outside into the cabin. Plastic tubes transmit tones to subject's ears holding the magnetic distortion to a minimum. The subject is instructed not to move, while the head is stabilized by air cushions in the dewar helmet of the multichannel MEG system. The pictures can be found at http://en.wikipedia.org/wiki/File:MSR_layered_door.jpg and http://infocenter.nimh.nih.gov/il/public_il/image_details.cfm?id=80.

[Häm93]. The cabin inside has to be demagnetized, while care has to be taken that only non-magnetic material is present in the cabin while recording.
At the level of MEG hardware, superconducting coil gradiometers are used. As the spatial gradient of the magnetic field of interest falls off rapidly, the field of interferers appear homogeneous in the surrounding of the human head. In contrast, fields from neuronal currents possess a high spatial gradient. Two radially oriented coils in opposite winding are used, one near the scalp and one a few centimeters away. Hence, homogeneous fields of external interferers are suppressed, while the recordings of interest are only weakly affected [Har05]. The current induced in the gradiometer is fed into a third coil producing a net magnetic field. This field, although less corrupted by external noise, is still weak. Super conducting quantum interference devices (SQUID) are used to record the weak magnetic fields with excellent sensitivity and low noise levels [Zim77]. Highly integrated sensors build the heart of today's MEG systems with up to 300 sensors cooled in a helium filled dewar. However, even if the dewar has a helmet shape, pickup coils are several centimeters away from the source. In the case of a current dipole source, the resulting magnetic field falls off as the inverse square of distance. Hence, it is important to instruct the subject to place his head as close as possible to the sensors, i.e. into the MEG helmet. If a dipole current is active in the brain, the map of the radial field component shows a dipolar pattern with two field extrema. The physical dipole is halfway between and at a right angle to a line connecting the

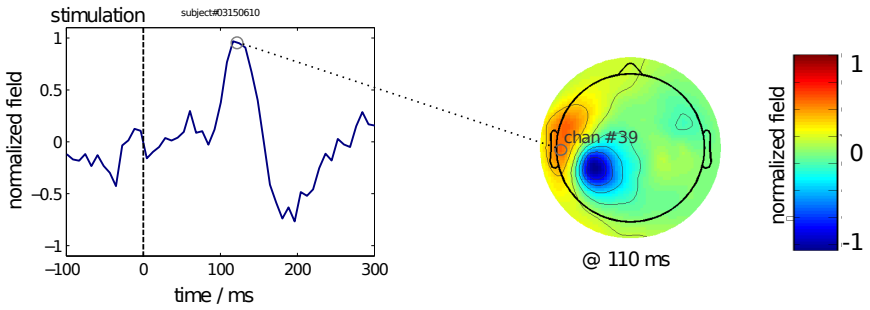


Figure 2.3: Event related field related to auditory stimulation. The depicted signal is the average over 100 single-trials, lined up to the stimulation time 0 ms. Single-trials were evoked by acoustic tones of 1 kHz and 30 ms duration, presented monaurally. The prominent deflection at 110 ms is referred to as N100. It is a typical peak that occurs after roughly 100 ms. The N100 varies from trial to trial and among subjects. Here, channel 39 was used to obtain the trial average, while the field map corresponds to the time instant 110 ms. Throughout this work, magnetic field plots are normalized to maximal field strength. The corresponding field maps are color coded from -1 to 1, indicating normalized field strength of the magnetic field vector pointing inside or outside the skull, respectively.

field extrema. This field can be approximated via quasi static Maxwell equations. Calculation of fields from a probe dipole is termed forward calculation. The resulting field is linearly related to the strength of the dipole and nonlinearly related to its position. In Fig. 2.3 a dipolar pattern of an auditory evoked source signal and the corresponding trial averaged response are displayed for one sensor.

2.2 Sensory evoked fields

Three sensory evoked fields are often used in a clinical setting [Wal05]: auditory evoked field (AEF), visual evoked field (VEF), and somatosensory evoked field (SEF). Parra and Walsh noted that stimulation paradigms lose clinical impact because of low specificity and better alternatives [Par03, Wal05]. However, Parra commented that a successful separation will let sensory evoked fields regain impact. Drug impacts, neuronal pathway integrity and diagnostic for epilepsy or multiple sclerosis are potential candidates in clinical applications [Nuw98]. For functional brain research, event-related paradigms are omnipresent and form a basis to understand many open questions [Mün00, Fra03].
Somatosensory fields are often evoked by electrical stimulation of peripheral nerves, AEFs are often evoked by tones and VEFs are often evoked by checkerboard-like patterns or flashlights. Throughout this thesis AEFs and VEFs shall be considered. To understand the experimental separation results, a brief neuroanatomical and physiological background is given, subsequently. For a review on somatosensory stimulation paradigms, refer to [Nie05].

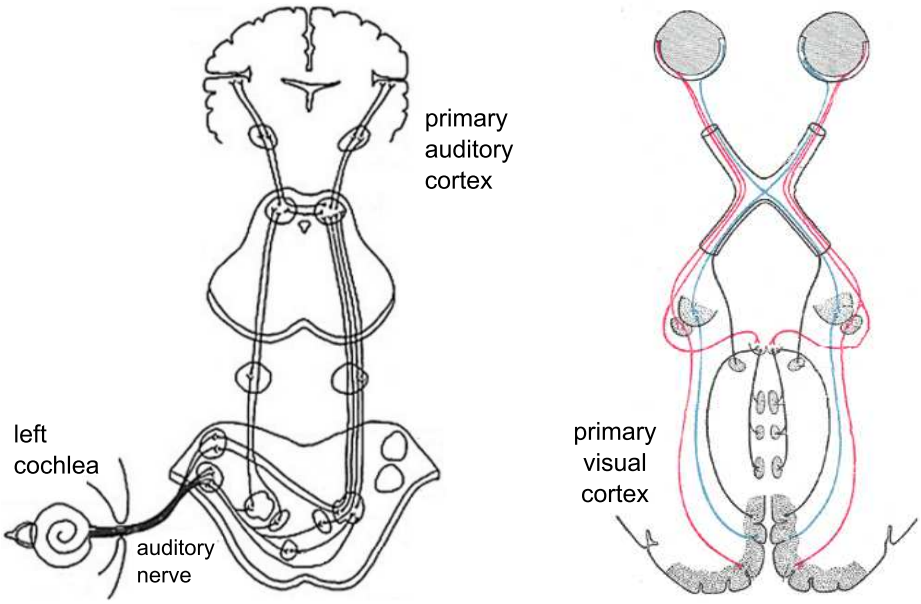


Figure 2.4: Left: Auditory pathway. After the signal leaves the cochlea, most of the nerves cross over to the contralateral side of the brain. However, some nerves carry on ipsilaterally. Hence, monaural stimulations, used in this work, evoke currents in the left and right primary auditory cortex yielding two dipolar field patterns, one over the left and one over the right temporal lobe.
Right: Visual pathway. The optic nerves from the nasal side of the retina cross at the optic chiasm. The optic nerves from the temporal side of the retina continue ipsilaterally. Hence, stimulations in the right visual field, used in this work, evoke currents in the left primary visual cortex yielding a single dipolar field pattern over the left occipital lobe. Pictures are modified versions of [Wil06, Gra18], respectively.

2.2.1 Auditory evoked fields

The primary auditory cortex is located in both hemispheres in the temporal lobe depicted in Figure 2.1. Acoustic stimulation results in electrical activity that travels along the auditory pathway, which is exemplified in Figure 2.4. The information is redistributed by many nuclei along the pathway to the primary auditory cortex. As a result, sound stimulations result in contralateral as well as ipsilateral neural activity; monaural as well as binaural stimulations give rise to field maps that have a dipolar pattern over both hemispheres. The magnetic response to acoustic stimulation has a prominent deflection after 90-110 ms, which is called the N100. The N100 was shown to vary with presentation conditions such as amplitude, spectrum or spatial sound direction [Nää87, Mäk88, Häm93]. Furthermore, different frequencies evoke different sites in the cortex [Pan88]. Changes in sounds [Mäk88] or sound sequences [Nää87] have been used extensively to study attention, reaction and memory and to study language disorders such as dyslexia. Due to single tone stimulation, two dipoles are assumed to generate the N100, one dipole in each hemisphere. However, more underlying sources (possibly less stable) may contribute [Lü92].

2.2.2 Visually evoked fields

The primary visual cortex is located in the occipital lobe depicted in Figure 2.1. In contrast to the auditory pathway, the visual nerves connect distinct parts of the retina to distinct parts of the primary visual cortex. Specifically, parts of the eyes that percept the left visual field are connected to the right primary visual cortex, whereas parts of the eyes that percept the right visual field are connected to the left primary visual cortex. This is exemplified in Figure 2.4. As a result, a dipolar magnetic field distribution appears contralateral to the stimulated visual field half.
Different types of visual stimuli exist. Commonly, pattern or flashlight stimuli are used with either full visual field or part visual field stimulation [Ain03]. An established paradigm is pattern reversal using checkerboard patterns [Ame94, Ahl92]. As with auditory evoked fields, the most prominent deflection is the visual P100, that peaks 90-140 ms after stimulation. A checkerboard stimulus reversing with 2 Hz was found to evoke a particularly large P100. However, the P100 was shown to vary widely with presentation condition such as dynamics, contrast, pattern orientation, spatial frequency and location [Sha00, Oka82, Ain03]. The P100 has been used to map the location of the visual fields to locations in the primary cortex [Slo99]. The number of regions activated is still not known. Recent studies give evidence that more than one region with possible similar temporal morphologies but larger time shifts are activated [Nää79, DaS91].

2.3 Neuronal basis and mathematical modeling

The neuronal basis of MEG data is of utmost importance for the design of separation methods, i.e. for the design of mathematical models.
On the one hand, a particularly advantageous physical property is the linearity of the mixing process. The field maps depend linearly on the amplitude of a spatial stationary source. They can be calculated using the quasi-static Maxwell equations [Sar87, Häm93].
On the other hand, any linear decomposition yields a set of results having certain features. Hence, knowledge about the biogenesis of neuronal currents and resulting magnetic fields yield prior information that is needed for exploratory data analysis. One may choose a model where only field maps with certain patterns are allowed, or one may choose a less restrictive model but interpret the results on the basis of some known features. If nothing is known about the genesis of the data, every separation method yields some separation results but conclusions cannot be drawn.

2.3.1 Currents and fields

By law of superposition, magnetic fields that result from opposite currents can cancel. Most of cell depolarization effects and resultant currents are of this kind and generate no external magnetic field [Zsc02]. In early days action potentials were believed to give rise to scalp potentials

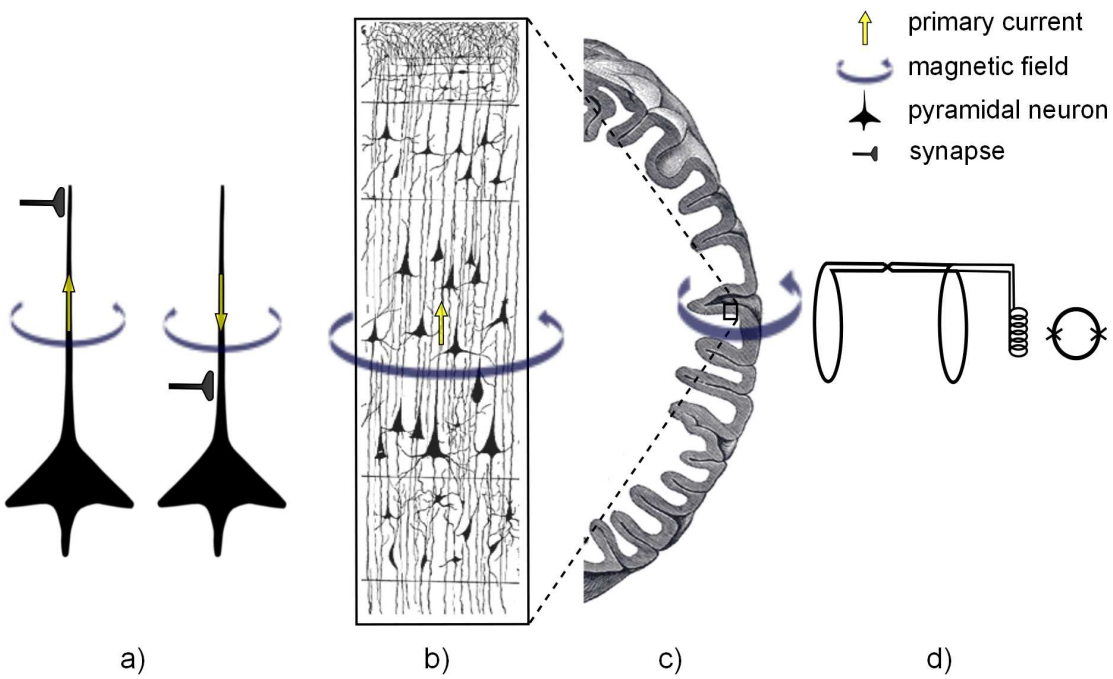


Figure 2.5: Electrophysiology, generation of magnetic fields and SQUID recording unit. The pictures show that currents can take both directions in the dendrites of long pyramidal cells (a). The latter are numerously arranged in parallel (b) and oriented perpendicular to the cerebral cortex (c). The pick-up coil registers the magnetic field which is fed to the SQUID sensor for recording (d).

[Zsc02]. However, a single action potential cannot be registered on the scalp. Furthermore, action potentials are wired in a chaotic fashion and, hence, canceling effects occur. From an MEG recording it is known that field variations occur smoothly and the spiky nature of action potentials therefore cannot be the generator of an MEG.
The main contributors to MEG are ionic currents in the dendrites of pyramidal cells [Zsc02]. This is possible due to a parallel arrangement of long pyramidal cells in the cortex. So called excitatory postsynapic potentials (EPSP) and inhibitory postsynapic potentials (IPSP) lead to depolarization or hyperpolarization of these cells, which results in intra and extracellular ionic currents. The extra-cellular currents (volume conduction) are seen in EEG, the fields of intracellular currents in MEG. For this, thousands of pyramidal cells have to be synchronously active. Postsynapic potentials last for several ms and new studies have shown that a single pyramidal cell can have up to 50000 synapses connected to it. Due to overlapping of many postsynapic potentials a rich variety of smooth signals can be generated sharing the variability of an MEG recording [Zsc02].
In Fig. 2.5 the biogensis effects are summarized. Fig. 2.5 a) shows that positive and negative currents can be generated, either by different locations of synapses or by different kind of synapses

(EPSP or IPSP). In b) the net flow in many pyramidal cells is shown to yield a net magnetic field that leaves and enters the cortex again. In MEG, this field is picked up with an axial gradiometer and measured with a SQUID sensor, as depicted in d) [Har05].

2.3.2 Linear instantaneous forward model

The forward model relates neural signal strength and magnetic field recorded by the MEG sensors. Hämäläinen [Häm93] showed that the human brain and skull are transparent to magnetic fields and that quasi static Maxwell equations apply. Hence, the forward model is linear and instantaneous. Sarvas [Sar87] showed that in a spherical approximation, secondary volume currents do not contribute to the magnetic field. Sinestra [Sti98] further noted that the volume current effect for MEG in more realistic shaped head models is very small. Haueisen [Hau02] noted that anisotropic volume conduction has a minor influence on source localizations but might have a major influence on source strength estimation.
Nevertheless, the most relevant primary current can be approximated using an equivalent current dipole (ECD) with a moment $\mathbf{Q}$ at the position $\mathbf{r}_0$. Using the isoconducting spherical head model, Sarvas solved the quasi-static Maxwell equations and the magnetic field $\mathbf{B}$ at the position $\mathbf{r}$ can be expressed as [Sar87]

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi F^2}(F\mathbf{Q} \times \mathbf{r}_0 - \mathbf{Q} \times \mathbf{r}_0 \cdot \mathbf{r}\nabla F), \tag{2.1}$$

where $F = |\mathbf{r} - \mathbf{r_0}|(|\mathbf{r}||\mathbf{r} - \mathbf{r_0}| + |\mathbf{r}|^{\mathbf{2}} - (\mathbf{r} \cdot \mathbf{r_0}))$. The model is nonlinear to the location and the orientation of the dipole. For a spatially fixed dipole, Eq. 2.1 shows that the forward model is instantaneous and linear in signal strength $|\mathbf{Q}|$. No field is observable, if the dipole is orientated radially; $\mathbf{Q} \times \mathbf{r}_0$ equals zero in the radial dipole setting.

2.3.3 Evoked versus phase-reset theory

There is a long standing debate regarding how evoked signals are generated [Nie09, Bec08, Mäk05, Kli06]. Basically, two theories are being discussed: the evoked model [Say74] and phase-reset model [Mak02]. The idea behind the evoked mechanism is that an event increases the signal power at some location in the brain, in addition to the ongoing processes. Thus, each stimulus activates neural populations in a time locked fashion. The phase-reset model refers to a view that stimulation changes ongoing activity nonlinearly by adjusting its phase domain after stimulation. ERFs seem to give evidence to the evoked mechanism. However, if the ongoing activity is not phase locked before stimulation, the same ERF is explainable by the phase-reset mechanism. Furthermore, there exist event-induced phenomena. The power of existent ongoing process was observed to increase or decrease with stimulation [Ber29]. Closing eyes blocks the alpha wave and generates beta waves, which is known as event-related (de-)synchronization [Pfu99]. Modeling the ongoing and the evoked mechanism together has been done recently in [Xu09].

2.3.4 Assumed evoked signal model

In this work event-related brain activity is modeled in simulations following the evoked model. Specifically, evoked MEG data is modeled in this work as a superposition of the magnetic fields that result from neuronal source signals. With each neuronal source signal, a particular spatial pattern of the magnetic field is associated. The single neuronal source signals and hence their magnetic fields are subject to amplitude and latency variations over trials. The evoked signals shape shall be modeled constant. Hence, MEG recorded data X_{cte} are assumed to follow the model expressed as

$$X_{cte} = \sum_f A_{cf} S_f(t + T_{ef}) D_{ef}, \tag{2.2}$$

where c, t, e, f denote channel, time instant, trial number and component number. The matrix $\mathbf{A}$ (with elements A_{cf}) encodes the spatial mapping from the neuronal source signals to the MEG channel, $\mathbf{T}$ the latency shifts and $\mathbf{D}$ the amplitude shifts. The fth neuronal signal form is denoted by $S_f(t)$. For notational convenience $S_{tf} = S_f(t)$ shall be used interchangeably. For instance, $A_{cf}S_{tf}$ is the magnetic signal at the cth channel evoked by the fth neuronal source at time t after stimulus presentation when no amplitude and latency variation is present.

2.4 Analysis methods

The analysis of evoked responses range from frequency analysis, Wiener filtering, time warping, wavelet filtering, statistical testing and component analysis of the ERP time domain signals. A good review on the univariate techniques to analyze and to enhance evoked signals can be found in [Pic95]. In the following, some of the considerations detailed in that work shall be discussed.

Averaging

The most widely used method for an analysis of evoked activity is the averaging method [Daw54], which uses a univariate model

$$x_e(t) = s(t) + n_e(t), \tag{2.3}$$

where $s(t)$ is the evoked signal.
The signal $s(t)$ is assumed to be invariant over $e = 1...E$ trials and the noise $n(t)$ is assumed white and additive.
The average over E trials is obtained via [Nie05]

$$x(t) = \frac{1}{E} \sum_e x_e(t), \tag{2.4}$$

while assuming $\mathrm{E}_e[x_e(t)] = s(t)$.
Since the noise is assumed zero mean, i.e. $\mathrm{E}_e[n_e(t)] = 0$, and the signal is assumed deterministic,

i.e. $\text{var}(s_e(t)) = \text{E}_e[(s_e(t) - \text{E}_e[s_e(t)])^2] = 0$ it follows from

$$\text{var}(x(t)) = \frac{1}{E}\text{var}(n(t)). \tag{2.5}$$

that the SNR improves linearly with the number of trials.
However, it is known that evoked signals can vary in amplitude and latencies [Tru02]. Furthermore, noise sources can be correlated and non-stationary with a certain degree of phase locking. Furthermore, Equation 2.3 is a univariate model; averaging does not use the multivariate nature of MEG nor does it model effects if more than one source is being evoked. Consequently, in reality, an improvement of SNR proportional to E is not achieved. Hence, averaging is a good technique to gain an overview over new data, but has to be regarded critically. In general, important single-trial details may be missed.

Localization

Localization aims at finding a physical location for associated activity. One method is to fit the forward calculated field of an ECD to a series of recorded field distributions in a least squares sense. Often the recordings are averaged over trials and, hence, only phase locked activity is localized. Furthermore, the averaged fields are hampered due to artifacts.
An alternative way is to use source separation as a preprocessing step prior to localization. Demixing yields signal vectors, while taking into account possible non-orthogonal noise sources. Furthermore each pattern is reduced to the number of dipoles that must be taken into account leading to a more plausible and faster result. As the inverse solution is non-unique, the result is always sensitive to the prior knowledge. Source separation, if successfully applied, can give clearer field maps, improving localization [Vig00, Tan02]. Studies showed that some sources are only detected and locatable when using ICA in a first step towards localization [Vig00, Tan02]. More detailed information on inverse methods can be found in [Bai01].

Independent component analysis

Decomposition makes use of the multivariate nature of the MEG recording. The aim is to find the underlying source signals, in contrast to the processing of superimposed single channel recordings. As the multi-channel recordings follow a linear instantaneous mixing process, a linear unmixing process has to be found.
Independent component analysis aims to recover univariate signals that underly the multivariate recordings. In the context of evoked signals, the aim is to obtain time dynamics and the associated field maps. Indeed, having the separated signals, improves single-trial analysis and the solution to the inverse problem. This makes ICA a vital tool for preprocessing in brain research [Tan05, Vig00, Par03, Ont06, Vig09]. The source signals, however, must be stationary and independent. The latter assumption will be under investigation in this work.

3 Independent component analysis

Independent component analysis (ICA) aims at recovering univariate source signals from multivariate recordings. The recorded data is assumed to be a mixture of these sources. Neither the source signal samples nor the mixing coefficients are known. This makes ICA an instance of the blind source separation (BSS) family. ICA assumes that all sources are stationary and statistically independent and that the recorded data is explainable by a linear mixing process. In the context of MEG data, the model underlying ICA is bilinear[1] - linear in the MEG recording modalities space and time.

3.1 Bilinear mixing model

A statistical instantaneous bilinear model may be expressed as

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{3.1}$$

where, in the context of ICA, $\mathbf{A} \in \mathbb{R}^{mxn}$ is called the mixing matrix. The random vector $\mathbf{s}$ comprises n source signals and the random vector $\mathbf{x}$ comprises the recorded data from m channels. The mixing matrix $\mathbf{A}$ is assumed to be stationary, square and of full rank.

ICA numerically adjusts an unmixing matrix $\mathbf{W}$ such that the estimates $\mathbf{u}$ are as independent as possible. An estimated source signal is one component of the recorded data and is recovered by left multiplication of the unmixing matrix with the data, given by

$$\mathbf{u} = \mathbf{W}\mathbf{x}. \tag{3.2}$$

In the case that $\mathbf{u} = \mathbf{s}$ holds true, ICA yields a clear copy of the underlying source signals by only using the mixture. Having MEG data of T time instances, Eq. 3.1 and Eq. 3.2 may be expressed in matrix notation as

$$\mathbf{X} = \mathbf{A}\mathbf{S}, \tag{3.3}$$
$$\mathbf{U} = \mathbf{W}\mathbf{X}, \tag{3.4}$$

where $\mathbf{X}$, $\mathbf{S}$, $\mathbf{U}$ represent the matrices of recorded data, source signals and recovered signals with T columns, respectively.

[1] Nonlinear, convolutive and noisy ICA models exist. These are often needed in more complex applications. For the MEG source separation task, the instantaneous linear model is used.

3.2 Identification

In order to identify a model, i.e. to estimate its parameters, the solution has to be existent and unique. In the case that ICA is based on the bilinear model, all parameters of the matrices $\mathbf{S}$ and $\mathbf{A}$ in Eq. 3.3 have to be estimated.
Let $\mathbf{A}_0 = \mathbf{AB}$ and $\mathbf{S}_0 = \mathbf{B}^{-1}\mathbf{S}$ be some matrices that differ from the original source matrix $\mathbf{S}$ and the original mixing matrix $\mathbf{A}$, due to multiplication by a regular matrix $\mathbf{B}$ and its inverse $\mathbf{B}^{-1}$. Let some recorded data following Eq. 3.3 be modeled by $\mathbf{A}_0$ and $\mathbf{S}_0$ yielding

$$\begin{aligned} \mathbf{X} &= \mathbf{A}_0\mathbf{S}_0 && (3.5)\\ &= \mathbf{ABB}^{-1}\mathbf{S}, && (3.6)\\ &= \mathbf{AS}, && (3.7) \end{aligned}$$

where $\mathbf{BB}^{-1}$ equals the identity matrix $\mathbf{I}$. Eq. 3.6 shows that infinitely many matrices $\mathbf{B}$ yield infinitely many candidate matrices $\mathbf{A}_0$ and $\mathbf{S}_0$ that all explain the recorded data equally well. The correct solution $\mathbf{AS}$ is not the only explanation of the data. As a consequence, the bilinear model is inherently not unique. In order to obtain uniqueness, constraints have to be imposed, i.e. the number of degrees of freedom have to be reduced.
Eq. 3.6 shall be further evaluated. Choosing $\mathbf{B}$ to be a regular permutation matrix $\mathbf{P}$ shows that a row change in the source signal matrix that is compensated by a corresponding row change in the mixing matrix destroys uniqueness. Choosing $\mathbf{B}$ to be a diagonal matrix $\mathbf{D}$ shows that energy can be assigned to either the mixing vector or to the associated source signal, which destroys uniqueness. Choosing $\mathbf{B}$ to be a signed identity matrix with different signs along its diagonal shows that changing the sign for both mixing vector and associated source signal destroys uniqueness. These indeterminacies, namely ordering, norm and sign, cannot be recovered without imposing constraints on the order, energy or sign of the source signals.
The solution to a constrained bilinear model, which is unique up to the indeterminacies ordering, norm and sign, shall be called *essentially unique*. An objective function, a *contrast* in the BSS context, that constrains some parameters of the bilinear model has to be imposed. The contrast must constrain the bilinear model, such that the solution is essentially unique.

3.2.1 PCA

Principal component analysis (PCA) [Str80, Gol96] is a frequently used signal processing technique. It has various applications with dimension reduction being its most prominent.
PCA may be introduced based on the bilinear model as stated in Eq. 3.1, in order to demonstrate its essential uniqueness. For this, PCA constrains the sources in the vector $\mathbf{s}$ to be uncorrelated as well as the mixing vectors to be mutually orthogonal. Then, Eq. 3.1 becomes

$$\mathbf{x} = \mathbf{Es}, \tag{3.8}$$

where $\mathbf{E}$ is orthogonal. Hence, the unmixing matrix $\mathbf{W}$ has to be orthogonal as $\mathbf{W} = \mathbf{E}^{-1} = \mathbf{E}^T$. The unmixing matrix is found utilizing the decorrelation constraint $\mathbf{C}_s = \mathrm{E}[\mathbf{ss}^T] \stackrel{!}{=} \mathbf{\Lambda}$, where $\mathbf{\Lambda}$

is diagonal. It follows with

$$\mathrm{E}[\mathbf{s}\mathbf{s}^T] = \mathrm{E}[\mathbf{E}^T\mathbf{x}\mathbf{x}^T\mathbf{E}] = \mathbf{E}^T\mathrm{E}[\mathbf{x}\mathbf{x}^T]\mathbf{E} = \mathbf{E}^T\mathbf{C}_x\mathbf{E} \stackrel{!}{=} \mathbf{\Lambda}, \tag{3.9}$$

that the data covariance matrix needs to be diagonalized. Hence, $\mathbf{E}$ and $\mathbf{\Lambda}$ can be found via an eigenvalue decomposition of the sample covariance matrix with $\mathbf{C}_x\mathbf{E} = \mathbf{E}\mathbf{\Lambda}$, where $\mathbf{E}$ is the eigenvector matrix and $\mathbf{\Lambda}$ is the eigenvalue matrix [Str80, Gol96].
Let $\mathbf{Q}$ be any non-diagonal orthogonal matrix and let $\mathbf{D}$ be any diagonal matrix. It follows with

$$\mathbf{Q}^T\mathbf{E}^T\mathbf{C}_x\mathbf{E}\mathbf{Q} = \mathbf{Q}^T\mathbf{E}^T\mathbf{E}\mathbf{\Lambda}\mathbf{E}^T\mathbf{E}\mathbf{Q} = \mathbf{Q}^T\mathbf{\Lambda}\mathbf{Q} \neq \mathbf{D}, \tag{3.10}$$

that $\mathbf{E}$ is essentially unique if all eigenvalues are distinct. However, if an eigenvalue appears repeated, the inequality in Eq. 3.10 turns into an equality, when deleting all columns and rows of $\mathbf{E}$, $\mathbf{Q}$, $\mathbf{\Lambda}$ and $\mathbf{D}$ that correspond to the distinct eigenvalues in $\mathbf{\Lambda}$, respectively. It follows that PCA is not essentially unique for eigenvectors that belong to the same eigenvalues [Hyv01a]. This has an impact on decomposition and preprocessing.

3.2.2 ICA

ICA is introduced based on the bilinear model as stated in Eq. 3.1, in order to demonstrate its essential uniqueness. For this, ICA constrains the sources in the vector $\mathbf{s}$ to be statistically independent. The mixing matrix is assumed square and regular. An ICA of the recorded data $\mathbf{x}$ aims at estimated sources $\mathbf{u} = \mathbf{W}\mathbf{x}$ that are as independent as possible. Comon proved that ICA is essentially unique, if $\mathbf{s}$ are pairwise mutually independent [Com94]. Furthermore, at most one source of $\mathbf{s}$ is allowed to be Gaussian distributed. This is a direct consequence of the Darmois-Skitovitch theorem [Dar53, Ski53], which may be expressed as [Com94]

Theorem 3.2.1. (Darmois-Skitovitch theorem) *Let s_i be independent random variables with $i = 1, ..., N$ and define two random variables as $x_1 = \sum_{i=1}^{N} a_i s_i$ and $x_2 = \sum_{i=1}^{N} b_i s_i$. Then if x_1 and x_2 are independent, all random variables s_j for which $a_j b_j \neq 0$ are Gaussian.*

ICA's independence constraint is more strict than PCA's decorrelation constraint. However, often, statistical independence is physically plausible. Furthermore, ICA drops the restriction to orthogonal mixing matrices; sources with non-orthogonal field maps can be separated by ICA.
Hence, although both PCA and ICA can be essentially unique, ICA may be regarded as superior in the decomposition context.

3.3 Framework of independence maximization

The most crucial assumption of ICA is that the sources are statistically independent. Independence is a stronger assumption than uncorrelatedness; independent sources must be uncorrelated, whereas uncorrelated sources may not be independent.
ICA is the framework under which different methods try to optimize an independence cost function. Theoretical basis and considerations shall be given in the following.

3.3.1 Definition of independence

Mutual statistical independence of two discrete random variables U and V is defined by

$$\mathrm{P}(U \in \mathrm{A}, V \in \mathrm{B}) = \mathrm{P}(U \in \mathrm{A})\mathrm{P}(V \in \mathrm{B}), \tag{3.11}$$

where A and B are any sets on two Cartesian coordinates and P denotes probability [Pap07]. For continuous random variables u and v, statistical independence is given if and only if the joint distribution $\mathrm{p}(u, v)$ factorizes into the product of the marginal distributions, given by

$$\mathrm{p}(u, v) = \mathrm{p}(u)\mathrm{p}(v). \tag{3.12}$$

ICA of mixed random sources that are pairwise mutually independent is essentially unique [Com94]. Hence, in the case that ICA recovers pairwise mutually independent estimates, these are guaranteed to be the sources if the ICA assumptions are met.
Let g and h be some absolutely integrable functions, it follows for the independent variables u and v that

$$\begin{aligned} \mathrm{E}[\mathrm{g}(u)\mathrm{h}(v)] &= \int_u \int_v \mathrm{g}(u)\mathrm{h}(v)\mathrm{p}(u, v)dudv, & (3.13)\\ &= \int_u \mathrm{g}(u)\mathrm{p}(u)du \int_v \mathrm{h}(v)\mathrm{p}(v)dv, & (3.14)\\ &= \mathrm{E}[\mathrm{g}(u)]\mathrm{E}[\mathrm{h}(v)] & (3.15) \end{aligned}$$

holds true [Hyv01a]. From Eq. 3.15 important properties of independence can be observed. Eq. 3.15 holds true also for non-independent uncorrelated random variables u and v, but if and only if g and h are linear functions. It follows that independence is a stronger constraint than uncorrelatedness.
However, if u and v are Gaussians, then uncorrelateness implies independence of u and v. This property does not hold true for distributions that differ from the Gaussian [Hyv01a]. Thus, after a decorrelation of two Gaussians, independence cannot assure essential uniqueness. Hence, no more than one Gaussian is allowed for ICA decomposition.

3.3.2 Whitening

ICA algorithms often iteratively adjust some cost function that is to be optimized over the parameter space. For large scale problems, this task is numerically challenging.
In the search of independent ICA sources, uncorrelateness of the sources is a necessary condition. By utilizing *whitening* as preprocessing, the recorded data are decorrelated and normalized to unit variance. The advantage is that the unmixing matrix is constrained to the manifold of orthogonal unmixing matrices [Car99, Hyv01a]. An increase in stability as well as numerical efficiency is obtained. Consequently, most ICA methods use whitening as a preprocessing step towards the goal of independent estimates.

Whitening is often performed by using the results from PCA. In particular, it follows from Eq. 3.9 that an eigenvalue decomposition of the data covariance matrix yield a matrix of eigenvectors that decorrelates the data. For normalization, element wise inversion of square root eigenvalues is used. The composite of the two matrices yields the whitening matrix

$$\mathbf{V} = \mathbf{\Lambda}^{-0.5}\mathbf{E}^T. \tag{3.16}$$

Applying Eq. 3.16 to the recorded data yields

$$\mathbf{x}_w = \mathbf{V}\mathbf{x} = \mathbf{V}\mathbf{A}\mathbf{s} = \mathbf{O}\mathbf{s}, \tag{3.17}$$

where $\mathbf{x}_w$ denotes whitened data. It can be observed from

$$\begin{aligned} \mathbf{C}_{\mathbf{x}_w} &= \mathrm{E}[\mathbf{x}_w\mathbf{x}_w^T] = \mathbf{V}\mathrm{E}[\mathbf{x}\mathbf{x}^T]\mathbf{V}^T && (3.18)\\ &= \mathbf{\Lambda}^{-0.5}\mathbf{E}^T\mathbf{E}\mathbf{\Lambda}\mathbf{E}^T\mathbf{E}\mathbf{\Lambda}^{-0.5} && (3.19)\\ &= \mathbf{\Lambda}^{-0.5}\mathbf{\Lambda}\mathbf{\Lambda}^{-0.5} && (3.20)\\ &= \mathbf{I}, && (3.21) \end{aligned}$$

that the elements of x_w are decorrelated and normalized to unit variance, i.e. white.
To see that the composite of whitening times mixing matrix $\mathbf{O} = \mathbf{V}\mathbf{A}$ is orthogonal the condition $\mathrm{E}[\mathbf{s}\mathbf{s}] = \mathbf{I}$ shall be employed. Diagonality is given by assuming independence; identity is obtained by rescaling. As a consequence, Eq. 3.18 may be rephrased as

$$\mathrm{E}[\mathbf{x}_w\mathbf{x}_w^T] = \mathbf{O}\mathrm{E}[\mathbf{s}\mathbf{s}]\mathbf{O}^T = \mathbf{O}\mathbf{I}\mathbf{O}^T \stackrel{!}{=} \mathbf{I}. \tag{3.22}$$

Hence, the composite matrix $\mathbf{O} = \mathbf{V}\mathbf{A}$ is orthogonal and so is the unmixing matrix[2] $\mathbf{W}_w = \mathbf{O}^{-1}$. The search space is reduced and numerical efficiency, stability, important information and theoretical results (cf. Sec. 3.3.4) can be gained.
Whitening is not essentially unique. From Eq. 3.18 and

$$\mathbf{Q}^T\mathbf{C}_{\mathbf{x}_w}\mathbf{Q} = \mathbf{Q}^T\mathbf{V}\mathrm{E}[\mathbf{x}\mathbf{x}^T]\mathbf{V}^T\mathbf{Q} = \mathbf{Q}^T\mathbf{I}\mathbf{Q} = \mathbf{I}, \tag{3.23}$$

it follows that any orthogonal transformation (or rotation) of the whitening matrix again yields data that is white. PCA suffers from an orthogonal (or rotational) invariance. This can be observed from the eigenvalue decomposition of $\mathbf{C}_{\mathbf{x}_w}$; all eigenvalues equal one and in line with Sec. 3.2.1 it follows that the associated eigenvector matrix is not essentially uniquely defined. Hence, ICA has to go beyond second order decorrelation in order to determine the remaining unknown rotation. The parameters of an orthogonal unmixing matrix are left to be estimated by ICA.

[2]The subscript indicates that the ICA unmixing matrix can be constrained orthogonal due to whitening. For convenience, the subscript will be omitted when the context is clear.

3.3.3 ICA via maximum likelihood

Maximum likelihood is a fundamental statistical technique for parameter estimation [Sch91]. It enjoys many favorable estimation theoretical properties, including consistency and asymptotically efficiency. Hence, it shall be used here as the first approach towards ICA.
The rule of strictly monotone density transformation relates the joint density of the random source vector to the joint density of the random data vector in Eq. 3.1 as [Pap07]

$$\mathrm{p}(\mathbf{x}) = \frac{1}{|\det\ \mathbf{A}|}\mathrm{p}(\mathbf{s}) = |\det \mathbf{W}|\mathrm{p}(\mathbf{W}\mathbf{x}), \tag{3.24}$$

where $\det$ denotes the determinant.
Using the assumption of source independence and inserting the rows of the unmixing matrix $\mathbf{w}_i$ into Eq. 3.24 yields [Car98a]

$$\mathrm{p}(\mathbf{x}) = |\det \mathbf{W}| \prod_{i=1}^{N} \mathrm{p}_i(\mathbf{w}_i\mathbf{x}), \tag{3.25}$$

where p_i are the marginal source densities. Consequently, the likelihood of T data samples as function of the parameters $\mathbf{W}$ is obtained as

$$\mathrm{L}(\mathbf{W}) = \prod_{t=1}^{T} |\det \mathbf{W}| \prod_{i=1}^{N} \mathrm{p}_i(\mathbf{w}_i\mathbf{x}_t), \tag{3.26}$$

and the log-likelihood follows as

$$\log \mathrm{L}(\mathbf{W}) = \sum_{t=1}^{T}\sum_{i=1}^{N} \log \mathrm{p}_i(\mathbf{w}_i\mathbf{x}_t) \ + \ T \ \log|\det \mathbf{W}|. \tag{3.27}$$

In order to perform maximum likelihood estimation, Eq. 3.27 has to be maximized with respect to its parameters. As the source densities are not parametrized, an infinite number of parameters are to be estimated. Clearly, this is problematic, given a limited number of observation samples. Cardoso demonstrated in [Car98a, Car03], that the exact source distribution is not crucial as long as the main characteristic of the underlying sources is respected. For now, let g be a nonlinear function somehow close to $\mathrm{g}_i = (\log \mathrm{p}_i)'$, where prime denotes the derivative.
The gradient may be expressed as [Hyv01a]

$$\frac{1}{T}\frac{\partial \log \mathrm{L}(\mathbf{W})}{\partial \mathbf{W}} = [\mathbf{W}^T]^{-1} + \mathrm{E}[\mathbf{g}(\mathbf{W}\mathbf{x})\mathbf{x}^T], \tag{3.28}$$

and a gradient based optimization given some step size α iteratively identifies the parameters of the unmixing matrix, given by

$$\mathbf{W}_{new} = \mathbf{W}_{old} + \alpha\frac{1}{T}\frac{\partial \log \mathrm{L}(\mathbf{W})}{\partial \mathbf{W}}\|_{\mathbf{W}=\mathbf{W}_{old}}. \tag{3.29}$$

From this brief ML ICA derivation, aspects that underly ICA by statistical independence maximization shall be discussed.

Eq. 3.24 has two factors, being the determinant of the unmixing matrix and the source joint distribution. The first accounts for a volume change due to the transformation of the source joint distribution. The latter holds information on the source joint distribution. Direct optimization of Eq. 3.24 puts a maximization pressure on the determinant term, which is maximal for an orthogonal unmixing matrix [Cho02]. Hence, an orthogonal mixing matrix is favored, which does not represent that the mixing matrix can be any regular matrix [Car98a].
For prewhitened data, the determinant is one and Eq. 3.24 amounts to finding the source joint distribution and the unmixing matrix. At this point, the rotational invariance becomes apparent. Once the source joint distribution has been found, any rotation of this joint distribution can be transformed to the data joint distribution equally via an orthogonal unmixing matrix, i.e. by counteracting the rotation. Hence, there is no point of absolute maximum of Eq. 3.24 over the parameter space but a line of equal solutions, representing non-uniqueness in the case that no constraint is imposed.
Eq. 3.25 ensures essential uniqueness and the pressure now is on the product of the marginal source densities [Cho02]. However, these have to be estimated from a limited number of samples. Specifically, exact independence measure involves exact estimation of densities. This is difficult and for ICA often more robust and more numerically efficient features of independence are in use. In this ML ICA derivation, an a priori fixed nonlinearity was used. Indeed, from Eq. 3.28, it can be observed that ICA is related to nonlinear decorrelation. To find a suited nonlinearity, hence, is an important aspect of ICA.
Eq. 3.29, finally, shows that the ICA parameters may have to be iteratively optimized. It is of importance which technique is used. Gradient ascent is not the best optimization technique as it is prone to slow convergence and inefficient use of information. More sophisticated techniques will be discussed later, which are detailed in [Gil81].

3.3.4 Information theoretical aspects

It was argued in the last section that evaluating statistical dependence is difficult. Many researchers have used information theoretical aspects to design and improve ICA objective functions. Hence, important concepts shall be discussed briefly. They will build the basis for the section on established ICA algorithms.
The *information* or *uncertainty* of an outcome of a discrete random variable x is defined as

$$\mathrm{H}(x_i) = \log \frac{1}{\mathrm{P}(x_i)} = -\log \mathrm{P}(x_i), \tag{3.30}$$

where P and $\log$ denote probability mass function and logarithm, respectively.
Originally introduced by Shannon [Sha48], entropy of a discrete random variable x equals the expected value of uncertainty, given by

$$\mathrm{H}(x) = \mathrm{E}[-\log \mathrm{P}(x_i)] = -\sum_i \mathrm{P}(x_i) \log \mathrm{P}(x_i). \tag{3.31}$$

The entropy of a continuous random variable is defined as

$$\mathrm{H}(x) = -\mathrm{E}[\log \mathrm{p}(x)] = -\int_x \mathrm{p}(x) \log\ \mathrm{p}(x) dx, \tag{3.32}$$

where p denotes the probability density function. Eq. 3.32 is sometimes called differential entropy [Cov91]. As $\log \mathrm{p}(x)$ is not defined for $\mathrm{p}(x) = 0$, the integration extends only over the support set, i.e. for $\mathrm{p}(x) \neq 0$. The seemingly trivial extension of the entropy concept from discrete-type to continuous-type random variables has limitations. Entropy given by Eq. 3.31 is always positive, whereas entropy defined as in Eq. 3.32 can become negative. As a consequence, properties of discrete-type entropy can differ from properties of continuous-type entropy. More details on the relation between discrete and continuous entropies can be found in [Cov91, Pap07, Köh05].
The entropy concept can be generalized to more than one random variable [Cov91]. For two random variables x and y, joint differential entropy is given by

$$\mathrm{H}(x, y) = -\mathrm{E}[\log \mathrm{p}(x, y)] = -\int_x \int_y \mathrm{p}(x, y) \log \mathrm{p}(x, y) dx dy, \tag{3.33}$$

while for a random vector $\mathbf{x}$ the joint differential entropy is given by

$$\mathrm{H}(\mathbf{x}) = -\mathrm{E}[\log \mathrm{p}(\mathbf{x})] = -\int_{\mathbf{x}} \mathrm{p}(\mathbf{x}) \log \mathrm{p}(\mathbf{x}) d\mathbf{x}. \tag{3.34}$$

Note that in ICA, the random vector $\mathbf{x}$ has elements that are linear mixtures of random source variables. An ICA adjusts the unmixing matrix $\mathbf{W}$ in order to minimize dependence of the estimated random vector $\mathbf{u} = \mathbf{W}\mathbf{x}$. For finding a suited objective function, here, dependency shall be discussed in terms of entropy.

In particular, Eq. 3.34 is a measure of joint uncertainty. Joint uncertainty can be different from marginal uncertainty as one element in $\mathbf{x}$ may already hold information about some other element. However, one element cannot increase the uncertainty of another element. Consequently, the inequality

$$\mathrm{H}(\mathbf{x}) \leq \sum_i \mathrm{H}(x_i) \tag{3.35}$$

must hold true. Eq. 3.35 states that if knowledge about one element in $\mathbf{x}$, say x_i, does not change the uncertainty of another element, say x_j, the random variables x_i and x_j are independent. In turn, if knowledge of x_i does change the uncertainty of x_j (and vice versa), these random variables are dependent. If uncertainty is not reduced by taking into account all elements of a random vector instead of one by one, equality in Eq. 3.35 holds true and the elements of the random vector are independent. If joint entropy is less than the sum of marginal entropies, some elements are dependent, i.e. hold information about each other.

Specifically, the information held by one random variable x_i about another x_j is termed as mutual information (MI), which results in

$$\mathrm{I}(\mathbf{x}) = \sum_i \mathrm{H}(x_i) - \mathrm{H}(\mathbf{x}). \tag{3.36}$$

Together with Eq. 3.35, and the above reasoning, mutual information is the measure of uncertainty reduction equaling the amount of statistical dependence of the random variables.
Inserting Eq. 3.34 into Eq. 3.36 yields

$$\mathrm{I}(\mathbf{x}) = \int_{\mathbf{x}} \mathrm{p}(\mathbf{x}) \log \frac{\mathrm{p}(\mathbf{x})}{\prod_i \mathrm{p}(x_i)} d\mathbf{x}. \tag{3.37}$$

This reformulation shows that mutual information is an instance of the more general Kullback-Leibler information measure, which is defined as

$$\mathrm{D}[\mathrm{p}(x)||\mathrm{q}(x)] = \int \mathrm{p}(x) \log \frac{\mathrm{p}(x)}{\mathrm{q}(x)} d\mathbf{x}. \tag{3.38}$$

The Kullback-Leibler distance quantifies differences between distributions; it is nonnegative, and zero only for equal distributions. It follows with

$$\mathrm{I}(x, y) = \mathrm{D}[\mathrm{p}(x, y)||\mathrm{p}(x)\mathrm{p}(y)] \geq 0 \tag{3.39}$$

that mutual information cannot become smaller than zero. If and only if $\mathrm{I} = 0$, the random variables are independent; in this case the joint distribution is assured to factorize into its marginals. However, optimization with mutual information as objective function is numerically difficult. Joint and marginal densities have to be estimated using a finite data sample. Thus, mutual information and related objective functions have to be elaborated further.
Evaluating Eq. 3.36 and 3.35 leads to important ICA concepts. Eq. 3.36 suggests that minimizing mutual information may be approached by minimizing marginal entropies or by maximizing joint entropy. The latter approach is a direct consequence of Eq. 3.35. The former approach is connected by the property of differential entropy of a Gaussian random variable. Assuming zero mean and fix variance, the Gaussian distribution has maximal entropy among all distributions [Cov91]. Furthermore, by the central limit theorem, a mixture of independent random variables approaches a random variable with a Gaussian distribution as the number of mixed variables grows [Pap07]. Hence, minimizing marginal entropies maximizes non-Gaussianity. In line with the central limit theorem, the most non-Gaussian elements must be the least mixed.
Together wit Eq. 3.39 and the central limit theorem, minimization of marginal entropies assures to lower dependence, i.e. to drive Eq. 3.36 to zero.
The numerical advantage is that only marginal distributions have to be considered. Although the estimation of marginal distributions remains challenging, these results provide a first step towards a more robust objective.
Differential entropy is not invariant to scaling of a random variable, i.e. it changes as

$$\mathrm{H}(ax) = \mathrm{H}(x) + \log |a|. \tag{3.40}$$

Hence, the scale has to be fixed *a priori* in order to determine differential entropy.
A scale invariant measure of non-Gaussianity is negentropy [Hyv01a], defined as

$$\mathrm{J}(\mathbf{x}) = \mathrm{H}(\mathbf{x}_G) - \mathrm{H}(\mathbf{x}), \tag{3.41}$$

where $\mathbf{x}_G$ has the same variance as $\mathbf{x}$, but it is defined to be Gaussian distributed. Due to the maximum entropy property of a Gaussian, negentropy is always greater than zero and zero if and only if the random vector $\mathbf{x}$ is Gaussian distributed.
If the recorded data is prewhitened, Eq. 3.36 can be reformulated as [Hyv01a, Köh00]

$$\mathrm{I}(\mathbf{x}_w) = \mathrm{J}(\mathbf{x}_w) - \sum_i \mathrm{J}((x_w)_i). \tag{3.42}$$

As negentropy is invariant to any linear transformation [Hyv01a], it follows that

$$-\frac{\partial}{\partial \mathbf{W}_w}\mathrm{I}(\mathbf{W}_w\mathbf{x}_w) = \frac{\partial}{\partial \mathbf{W}_w}\sum_i \mathrm{J}((\mathbf{W}_w\mathbf{x}_w)_i). \tag{3.43}$$

which proves that, in the case of prewhitened data, minimization of marginal entropies is equivalent to minimization of mutual information.
This is an important finding, stating that the concept of non-Gaussianity is applicable for ICA. In this context, the 4th order cumulant *kurtosis* can be used advantageously. Kurtosis $\kappa(x)$ is a higher order statistic of a random variable, which is zero for a Gaussian distributed signal; it is negative for sub-Gaussian (less peaky) distributed variables and positive for super-Gaussian (more peaky) distributed variables [Hyv01a]. Given a zero-mean variable, kurtosis $\kappa(x)$ is calculated as

$$\kappa(x) = \mathrm{E}[x^4] \; - \; 3\mathrm{E}[x^2]^2. \tag{3.44}$$

The relation of maximal non-Gaussianity and minimal mutual information, together with making use of kurtosis, yields the first practical approach towards ICA without an estimation of full underlying source distributions. However, kurtosis is known to be corrupted by outliers [Hyv01a] and may not be fully robust for ICA in practice. In the next section, established ICA methods are presented; most of them draw directly on mutual information and its properties. In particular, this section provides the basis for more efficient and more robust ICA methods that will be presented subsequently.

3.3.5 Quality measure

A normalized version of Amari's performance index [Ama96] quantifies the unmixing error err, which may be expressed as

$$\mathrm{err}(\mathbf{G}) = \frac{1}{2m(m-1)}\left[\sum_j\left(\sum_i \frac{|g_{ij}|}{\max_k|g_{kj}|} - 1\right) + \sum_i\left(\sum_j \frac{|g_{ij}|}{\max_k|g_{ik}|} - 1\right)\right], \tag{3.45}$$

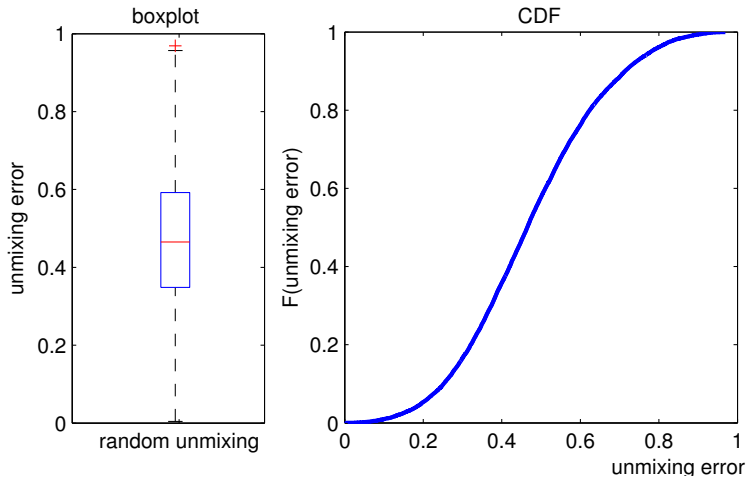


Figure 3.1: The unmixing error err will be used throughout this thesis as a measure of separation success. The unmixing error is a variant of Amari's performance index, which measures the deviation from a permuted identity matrix. It ranges between zero and one and is zero only if unmixing is perfect. In the figure, random 2×2 dimensional mixing matrices $\mathbf{A}$ and unmixing matrices $\mathbf{W}$ were generated. For each draw, each parameter of each matrix was a sample of a uniform distribution in [-1,1]. Hence, the figure provides statistics of the unmixing error err in the case of random unmixing.

where the $m \times m$ matrix $\mathbf{G} = \mathbf{WA}$ is the product of the estimated unmixing matrix $\mathbf{W}$ with the *a priori* known mixing matrix $\mathbf{A}$. The unmixing error err is always between 0 and 1 and equal to 0 if and only if $\mathbf{G} = \mathbf{PD}$, where $\mathbf{P}$ is a permutation matrix and $\mathbf{D}$ is a diagonal matrix. Hence, at this point, unmixing is perfect up to the indeterminacies permutation, sign and norm. For prewhitened data $\mathbf{G}$ equals $\mathbf{W}_w\mathbf{VA}$, where $\mathbf{V}$ is the whitening matrix and $\mathbf{W}_w$ is the unmixing matrix recovered from whitened data.

3.4 Established methods

3.4.1 FastICA

FastICA aims at minimal dependence among the components $\mathbf{u}$ of the data vector $\mathbf{x}$ by minimizing the marginal entropy of each component. In Section 3.3.4, minimizing marginal entropy was shown to be one approach towards minimizing mutual information. When the components are constrained to have zero mean and fixed variance, minimization of marginal entropy can be obtained by maximization of non-Gaussianty of each component.
Negentropy was introduced in Eq. 3.41 and shown to be a scale invariant measure for non-Gaussianity. Using prewhitening, due to Eq. 3.41 to 3.43, maximization of negentropy was shown to drive the components towards independence. Negentropy can be approximated by a truncated

Gram-Charlier expansion [Com94] as

$$\mathrm{J}(x) \approx \frac{1}{12}\mathrm{E}[x^3]^2 + \frac{1}{48}\kappa(x)^2. \tag{3.46}$$

For symmetrically distributed signals, the first term vanishes. Thus, the maximum of Eq. 3.46 is found by maximizing the absolute value of kurtosis $\kappa(x)$.
Given whitened data[3], the gradient of kurtosis may be expressed as

$$\frac{\partial}{\partial \mathbf{w}}|\kappa(\mathbf{w}^T\mathbf{x})| \propto \text{sign }(\kappa(\mathbf{w}^T\mathbf{x}))(\mathrm{E}[\mathbf{x}(\mathbf{w}^T\mathbf{x})^3]). \tag{3.47}$$

Simple gradient ascent is obtained via

$$\mathbf{w}_{new} \quad = \quad \mathbf{w}_{old} + \mu\frac{\partial}{\partial \mathbf{w}}|\kappa(\mathbf{w}^T\mathbf{x})|_{\mathbf{w}=\mathbf{w}_{old}}. \tag{3.48}$$

The norm of the unmixing vectors can be set to unity after each step as $\mathbf{w}_{new} = \mathbf{w}_{new}/||\mathbf{w}_{new}||$ [Hyv01a]. This normalization technique is often used in ICA methods, as the scale remains undetermined by ICA. It is the direction of the unmixing vector $\mathbf{w}_i$ only, that unmixes the component.
The kurtosis objective function grows fast due to the 4th power in Eq. 3.44. Thus, optimization of the absolute kurtosis value is not robust against outliers. Indeed, outliers can strongly dominate the estimated value of kurtosis, which can destroy separation [Hyv01a].
Negentropy is based on the maximum entropy property of a Gaussian. The key property for a maximum entropy objective function is being nonnegative and zero if and only if the component is Gaussian distributed. Hence, Eq. 3.46 represents *one* approximation and may be generalized to a family of approximations, given by [Hyv01a]

$$\mathrm{J}(x) \approx (\mathrm{E}[\mathrm{G}(x)] - \mathrm{E}[\mathrm{G}(x_{\text{Gauss}})])^2, \tag{3.49}$$

where G is some nonlinearity that is non-quadratic. Then, Eq. 3.49 represents a general measure of non-Guassianity, being nonnegative and zero if and only if x is a Gaussian.
In particular, a more robust measure of non-Guassianity than kurtosis is obtained, when G grows less fast than to the 4th power. Hyvärinen tested various nonlinearities, among which the nonlinearity

$$\mathrm{G}(x) = \frac{1}{a}\log\cosh ax, \tag{3.50}$$

where $1 \leq a \leq 2$, proved more suitable than kurtosis in many practical cases [Hyv97].
Hyvärinen proposed to optimize Eq. 3.49 by a fixed point algorithm rather than by plain gradient-based optimization, for the sake of numerical robustness and efficiency. He showed that the convergence improved significantly. The corresponding update rule is given by [Hyv97]

$$\mathbf{w}_{new} \quad = \quad \mathrm{E}[\mathbf{x}\mathrm{G'}(\mathbf{w}_{old}^T\mathbf{x})] - \mathrm{E}[\mathrm{G''}(\mathbf{w}_{old}^T\mathbf{x})]\mathbf{w}_{old}, \tag{3.51}$$

[3] In this section, all data vectors are assumed white. The whitening index w will be ommited in this section for notational convenience.

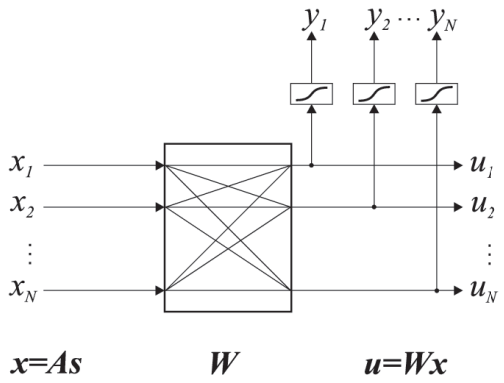


Figure 3.2: Infomax transforms the ICA components $\mathbf{u}$ by using nonlinearities. Maximal joint entropy of the transformed components $\mathbf{y}$ assures that its elements are independent and distributed uniformly. The picture is taken from [Köh05].

where G′ denotes the first derivative of Eq. 3.50, given by

$$\mathrm{G}'(x) = \tanh ax \tag{3.52}$$

and G″ denotes the second derivative of Eq. 3.50, given by

$$\mathrm{G}''(x) = a(1 - \tanh^2 ax). \tag{3.53}$$

Again, normalization of the ummixing vectors is imposed after each iteration step.
In summary, a modified version of negentropy ensures to maximize non-Gaussianity, while choosing some other nonlinearity than kurtosis yields a more robust negentropy measure. Going beyond simple gradient-based optimization yields more efficiency and higher convergence rates. Maximization of non-Gaussianity, in turn, minimizes marginal entropy and dependence of the estimated ICA components. The result is a robust and fast ICA algorithm, referred to as FastICA.

3.4.2 Infomax

Infomax approaches independence by maximization of joint entropy. However, even though joint entropy is always smaller than the sum of the marginal entropies (Eq. 3.35), maximization of joint entropy does also put a maximization pressure on the marginal entropies, which results from Eq. 3.36. Furthermore, the components are zero mean and unit variance and hence, maximal marginal entropy is obtained when the components are maximally mixed! This is a direct consequence of the central limit theorem and the maximal entropy property of a Gaussian.
However, applying nonlinear transformation units

$$y = \mathrm{g}(\mathbf{w}^T\mathbf{x}) = \mathrm{g}(u), \tag{3.54}$$

to each estimated component makes the maximum joint entropy concept applicable. The transformation stage is depicted in Fig. 3.2. To see the validity, let a nonlinearity of the type

$$\mathrm{g}: \mathrm{R} \to (0;1), \qquad \text{strictly monotonous} \tag{3.55}$$

bound the amplitudes of the transformed signal y. In this case, marginal entropies become maximal if each transformed component is uniformly distributed. With the rule of density transformations [Pap07], marginal entropy is given by

$$\begin{aligned} \mathrm{H}(y) &= -\mathrm{E}[\log \mathrm{p}(y)], && (3.56)\\ &= -\mathrm{E}\left[\log \frac{\mathrm{p}(u)}{\partial y/\partial u}\right], && (3.57)\\ &= -\mathrm{E}\left[\log \frac{\mathrm{p}(u)}{\mathrm{g}'(u)}\right], && (3.58)\\ &= -\mathrm{D}[\mathrm{g}'(u)||\mathrm{p}(u)] \leq 0. && (3.59) \end{aligned}$$

Due to the Kullback Leibler distance in the last identity, marginal entropy $H(y)$ becomes maximal if the derivative g' equals the distribution $\mathrm{p}(u)$. As g is obtained by integration of the derivative g', the nonlinearity g must be the cumulative distribution function of the component u. Indeed, transformation of a random variable by its own cumulative distribution function yields a uniformly distributed random variable [Pap07]. Hence, maximization of the joint entropy

$$\mathrm{H}(\mathbf{y}) = \sum_i \mathrm{H}(\mathrm{g}_i(u_i)) - \mathrm{I}(\mathbf{y}) \tag{3.60}$$

under the constraint $g' = \mathrm{p}(u)$ leads inherently to maximal marginal entropies. Further maximization is only obtainable by reducing the dependence between the components driving these to independence. Hence, maximization of joint entropy of the transformed vector $\mathbf{y}$ yields independence of the estimated components.
For numerical implementation, let

$$\mathbf{y} = (y_1(u_1), ..., y_N(u_N)) = (\mathrm{g}_1(w_1^T\mathbf{x}), ..., \mathrm{g}_N(w_N^T\mathbf{x})). \tag{3.61}$$

Substituting Eq. 3.61 into Eq. 3.60 yields

$$\begin{aligned} \mathrm{H}(\mathbf{y}) &= -\mathrm{E}[\log \mathrm{p}(\mathbf{y})], && (3.62)\\ &= -\mathrm{E}\left[\log \frac{\mathrm{p}(\mathbf{x})}{|\det \partial\mathbf{y}/\partial\mathbf{x}|}\right] && (3.63)\\ &= \mathrm{H}(\mathbf{x}) + \sum_i \mathrm{E}[\log \mathrm{g}_i'(\mathbf{w}_i^T\mathbf{x})] + \log|\det \mathbf{W}|, && (3.64) \end{aligned}$$

which is equal to the expression stated in Eq. 3.27. As a result, Infomax and maximum likelihood are equivalent methods. Both have to estimate the source densities from a finite sample size. Infomax uses a fixed source density, the logistic nonlinearity

$$y = \mathrm{g}(u) = \frac{1}{1 + e^{-u}}. \tag{3.65}$$

Together with Eq. 3.65, the Infomax update rule may be expressed as [Bel95]

$$\begin{aligned} \mathbf{W}_{new} &= \mathbf{W}_{old} + \alpha(\mathbf{W}_{old}^{-T} + \frac{\partial}{\partial \mathbf{W}} \sum_i \mathrm{E}[\log \mathrm{g}_i'(\mathbf{w}_i^T \mathbf{x})]), & (3.66) \\ &= \mathbf{W}_{old} + \alpha(\mathbf{W}_{old}^{-T} + (\mathbf{1} - 2\mathbf{y})\mathbf{x}^T). & (3.67) \end{aligned}$$

Utilizing the natural gradient [Ama98] by right multiplying Eq. 3.67 with $\mathbf{W}^T\mathbf{W}$, the speed of convergence can be significantly improved in contrast to standard gradient techniques.

3.4.3 JADE

JADE generalizes the idea of FastICA by evaluating fourth-order crosscumulants as well as the autocumulant kurtosis. Specifically, JADE aims at diagonalizing a four-way tensor holding all fourth order cumulant statistics. This approach is valid, as crosscumulants of independent signals are zero [Nik93].
A fourth-order cumulant tensor may be represented by $N \times N$-dimensional cumulant slices [Car93], defined as

$$n_{ij} = \sum_{k,l=1}^{N} \mathrm{cum}(x_i, x_j, x_k, x_l) m_{kl}, \tag{3.68}$$

where the matrix $\mathbf{M}$ with elements m_{kl} has a single non-zero element equal to one. There are N^2 different matrices M and hence a set of N^2 cumulant slices $\mathbf{N}$ exist. Using prewhitened data constrains the matrix $\mathbf{O} = \mathbf{VA}$ to be unitary. Using the cumulative as well as the multi-linearity property of cumulants [Nik93], Eq. 3.68 may be expressed as

$$n_{ij} = \sum_{f=1}^{N} \sum_{k,l=1}^{N} o_{if} o_{jf} o_{kf} o_{lf} \mathrm{cum}(s_f, s_f, s_f, s_f) m_{kl}, \tag{3.69}$$

where $\mathrm{cum}(s_f, s_f, s_f, s_f)$ equals kurtosis of the fth signal. In matrix notation, Eq. 3.69 becomes [Car93]

$$\mathbf{N} = \sum_{f=1}^{N} \kappa(s_f)(\mathbf{o}_f^T \mathbf{M} \mathbf{o}_f)\mathbf{o}_f \mathbf{o}_f^T = \sum_{f=1}^{N} \lambda_f \mathbf{o}_f \mathbf{o}_f^T. \tag{3.70}$$

Eq. 3.70 represents an eigenvalue problem and $\mathbf{O}$ can be found via an EVD of $\mathbf{N}$. A more robust estimate of $\mathbf{O}$ is obtained when jointly diagonalizing the whole set of N^2 cumulant matrices by

joint Jacobi transformation [Car96b].
Furthermore, Cardoso showed that for the set of matrices $\mathbf{M}$, N matrices can fully exhaust the fourth order information. Specifically, expressing Eq. 3.68 as $\mathbf{N}(\mathbf{M}) = \lambda\mathbf{M}$ yields N matrices in $\mathbf{M}$ with Eigenvalues greater than zero [Car93]. Joint approximate diagonalization of a reduced set of only significant cumulant matrices amounts to a robust and efficient ICA method. Robustness is obtained at the price of evaluating a high number of fourth-order cumulant statistics. Numerical efficiency is obtained by matrix algebra-based simplifications.

3.4.4 SOBI

Second order blind identification (SOBI) [Bel97] as well as time domain separation (TDSEP) [Zie98] use time delayed second order statistics. If the underlying sources are random variables without temporal structure, all second order information is exhausted by prewhitening. As a consequence, separation is not possible based on second order statistics. However, if the sources have spectral color, then further second order information can be used. If two components are decorrelated for all time shifts τ, which may be expressed as

$$\mathrm{E}[s_i(t)s_j(t+\tau)] = 0, \tag{3.71}$$

blind source separation based on second order statistics is possible.
Different time lags of the recorded signals yield different covariance matrices. Using prewhitened data, these covariance matrices are calculated as

$$\mathbf{R}_i = \mathrm{E}[\mathbf{x}(t)\mathbf{x}^T(t+\tau_i)] = \mathbf{O}\mathrm{E}[\mathbf{s}(t)\mathbf{s}^T(t+\tau_i)]\mathbf{O}^T = \mathbf{O}\mathbf{\Lambda}_i\mathbf{O}^T. \tag{3.72}$$

Based on two covariance matrices from two time lags τ_i and τ_j, the solution to the generalized eigenvalue problem [Mol94]

$$\mathbf{O}^T(\mathbf{R}_j^{-1}\mathbf{R}_i) = (\mathbf{\Lambda}_j^{-1}\mathbf{\Lambda}_i)\mathbf{O}^T \tag{3.73}$$

yields the whitening mixing composite matrix $\mathbf{O} = \mathbf{VA}$ and gives the unmixing matrix $\mathbf{W} = \mathbf{O}^T$. Joint Jacobi diagonalization [Car96b] of a set of time lagged covariance matrices given by Eq. 3.72 yields more robust estimates.
The advantage of SOBI is that also (colored) Gaussian sources can be separated. The disadvantage is that source signals need to have different spectral profiles. This can be seen from Eq. 3.72. If the autocorrelation of the ith signal at lag τ is equal to the autocorrelation of the jth signal at lag τ than the eigenvalues are equal. Hence, the eigenvectors are not essentially uniquely defined and these signals are not separable at this time lag. When the spectra are equal then no time lag can be found that makes the autocorrelation distinct. Hence, for similar spectra, SOBI loses performance. Likewise, for overlapping spectra the right choice of a set of τ_i becomes difficult [Hyv01a].

4 On ICA for evoked dependent source signals

The motivation for using ICA for the task of source separation is its elegance. ICA does not assume knowledge of the physical system but blindly infers information about both the mixing process and the sources. Often, independence is a natural property of these sources and linear mixing can be assumed or more sophisticated models, such as the convolutive, are used.
Nevertheless, real world signals are often not perfectly independent. If they are, any finite sample realization gives rise to estimation errors. Indeed, many researchers have stressed that ICA yield components that are *as independent as possible*. This is true in the sense that an ICA method always favors the objective it is designed for, i.e. to maximize independence of the recorded components. However, using the bilinear model, linear combination of the data will always yield some local maximum of an independence objective function, even in the case when the signals are dependent!
The question to ask is whether the optimization criterion allows the assessment of the independence assumptions of the sources. Clearly, in the end, relevant is *how close* the estimates are from the truth and not *how independent* the estimates are after an ICA.
In this chapter, answers in the context of evoked MEG sources are given. For this, aspects based on theory as well as virtual and actual data assessment will be presented.

4.1 Statistical dependence

4.1.1 Theory

Statistical independence of random sources equals zero Kullback-Leibler distance between the joint distribution of random variables and the product of corresponding marginal distributions. This distance was termed as mutual information, in Sec. 3.3.4, Eq. 3.37. In order to discuss statistical dependence in detail, Eq. 3.37 shall be repeated for two discrete random variables[1], given by

$$\mathrm{I}(x,y) = \sum_i \sum_j \mathrm{P}(x_i, y_j) \log \left(\frac{\mathrm{P}(x_i, y_j)}{\mathrm{P}(x_i)\mathrm{P}(y_j)} \right). \tag{4.1}$$

[1] Note that real world signals are usually not represented by fine alphabet. However, sampled data signals always are. Hence, the discussion of mutual information in terms of Eq. 4.1 is closely related to the mechanism of ICA in practice.

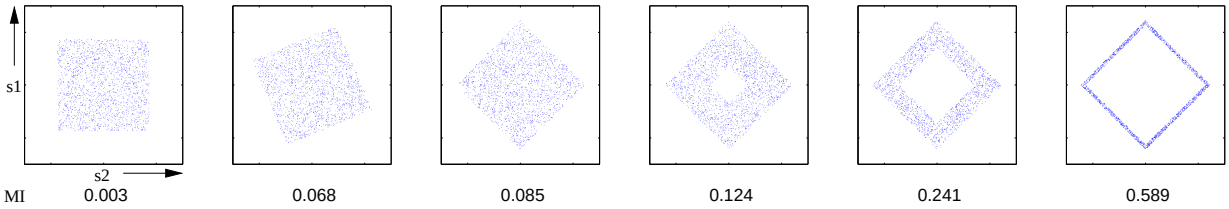


Figure 4.1: Mutual information of six joint distributions ranges from 0.003 to 0.589, i.e. the depicted signal pairs have no, little and strong dependencies. These were generated by adjusting the original joint of the two signals s_1 and s_2 with uniform marginal distributions, via rotation and erosion.

A thought experiment shall support deeper understanding of Eq. 4.1. Let x and y be two discrete random sources having finite alphabet, each. The two random variables are independent if and only if $\mathrm{P}(x_i, y_j)$ and $\mathrm{P}(x_i) \cdot \mathrm{P}(y_j)$ are equal for every combination of indices i and j in Eq. 4.1. For a more detailed discussion, let x and y be binary, i.e. $x, y \in \{0, 1\}$. Now, infinitively many samples of the binary random variables x and y shall be drawn, in tuples. Each variable can take on only two different values, 0 or 1, whereas four different tuples (x, y) *may* occur, namely $(0, 0)$, $(0, 1)$, $(1, 0)$ or $(1, 1)$. If they all occur uniformly distributed, then $\mathrm{P}(x_i, x_j)$ equals 0.25 for all combinations of indices i and j, whereas $\mathrm{P}(x_i)$ and $\mathrm{P}(x_j)$ equal 0.5 for $i = 1, 2$ and $j = 1, 2$. Hence, for all combinations of indices i and j, $\mathrm{P}(x_i, y_j)$ equals $\mathrm{P}(x_i) \cdot \mathrm{P}(y_j)$. Thus, the two random variables are independent and, consequently, the logarithm in Eq. 4.1 is always zero. That is, knowing the value of one source does not give information about the value of the other source and the mutual information is zero.
In contrast, if knowledge of x implies the value of y, i.e. the entropy of y is reduced to zero when x is known, then only the tuples $(0, 1), (1, 0)$ or the tuples $(0, 0), (1, 1)$ occur. It follows that no combination of i and j result in $\mathrm{P}(x_i, y_j) = \mathrm{P}(x_i) \cdot \mathrm{P}(y_j)$ and highest dependence is obtained. Note that the occurrence of just one single tuple, say $(1, 1)$ does not result in an even higher dependence. In this case, the corresponding sources are no longer random and entropies and mutual information are zero.
An intermediate dependence level is achievable, when next to two tuples, say $(0, 1), (1, 0)$, also one of the remaining tuples, i.e. $(1, 1)$ or $(0, 0)$, occurs. For the latter, it is known that y must be zero when x is one, but nothing is known about the value of y in the case that x is zero. Evaluating Eq. 4.1 for these scenarios, assuming uniform distributions, mutual information equals 0 for the independent scenario, $0,69$ for the dependent scenario and $0,29$ for the intermediate dependent scenario.
From these elaborations, it can be concluded that if certain sample tuples have different probability than their marginal distributions would suggest, then these signals are dependent. If the knowledge of one value from one source make certain values of the other source more probable, then dependencies occur. The degree of dependence can be measured by mutual information, given by Eq. 4.1.
Dependence can also be analyzed visually regarding the joint distribution of point tuples of two signals. Fig. 4.1 depicts six joint distributions, each belonging to two discrete random sources.

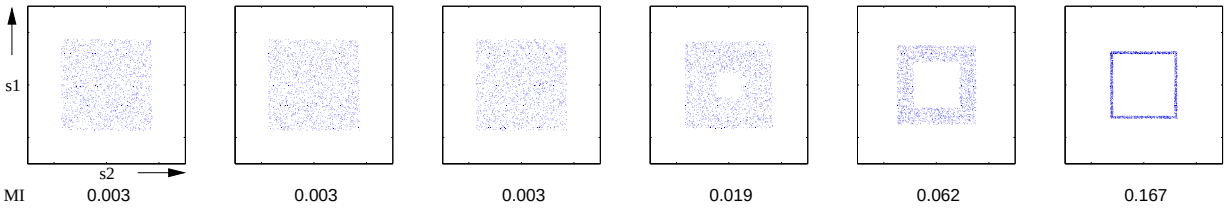


Figure 4.2: Each signal pair of Fig. 4.1 is directly fed to ICA without any mixing. Here, the outcomes are depicted, respectively. The resulting joint distributions suggests that all but the first has been changed by ICA. Indeed, dependencies are reduced. However, the last 3 joint distributions still show considerable amount of dependencies, that may not be reduced by any linear transformation.

The first graph in Fig. 4.1 depicts the joint distribution of two independent signals s_1 and s_2 with uniform marginal distributions. Indeed, regarding a fixed value of s_1 does not say anything about the value of s_2. Hence, every sample tuple can occur with the same probability being the product of the marginal distributions.
The third graph in Fig. 4.1 depicts a joint distribution that is a 45 degree rotated version of the first. Here not every sample tuple can occur. The max of s_1 forces s_2 always to be zero. Hence, knowledge of s_1 reduces the entropy of s_2. The entropy can be reduced further by reducing the number of tuples that can occur. Taking points from the joint distribution by cutting a square with increasing edge length, yields more and more dependent signals. The mutual information value is given below each joint distribution depicted in Fig. 4.1. Specifically, mutual information ranges from 0.003 to 0.589 and, thus, six signal pairs with almost no, little and strong dependencies are generated by adjusting the original joint distribution of the two signals s_1 and s_2 with uniform marginal distributions, accordingly.

4.1.2 Why study ICA in such a context?

ICA is based on the concept of minimizing statistical dependence. If the signals to be separated are known or conjectured to be dependent, why should ICA be used? Why study ICA in such a context?
Indeed, dependence tailored algorithms may appear to be more appropriate. However, there is only one independent joint distribution and unlimited dependent joint distributions of two or more sources. This makes the choice of a tailored dependency a difficult one.
Furthermore, connectivity [Nol04] must not be confused with statistical independence. Two signals that are shifted copies have high connectivity; their cross-correlation at the shift lag is high. However, if no temporal overlap exists, mutual information is low! Consequently, two evoked signals may be independent, although being connected to stimulation. Hence, it is of utmost importance to investigate the existence and the influence of dependencies on the performance of ICA in order to correctly design and interpret experiments for evoked brain research.

4.1.3 Can ICA separate dependent source signals?

Can ICA separate dependent signals? This question seems trivial. If ICA assumes independence but the underlying signals are dependent, ICA does not seem to be suitable. Nevertheless, some general ideas on how ICA act on dependent sources shall be considered here.
Let the six signal pairs that are depicted in Fig. 4.1 and Fig. 4.2 be the material for another experiment. These signal pairs in Fig. 4.1 have almost no, little and up to high dependencies. Each signal pair shall now directly be fed to ICA without any mixing. Fig. 4.2 shows the joint distributions of the ICA estimates. These suggest that all but the first joint distribution has been changed by ICA. Comparing the mutual information of the joint distributions depicted in Fig. 4.1 and depicted in Fig. 4.2, shows that ICA has been successful in reducing dependencies. The three latter joint distributions still have dependencies and these components are only 'as independent as possible'.
However, successful is the right term only if the task is to reduce dependency. In the described experiment, ICA was applied to sources that have not been mixed. Instead, all the signal pairs (but the first) were *mixed* in order to achieve a dependency reduction under a linear transformation. If ICA was successful in separation, the joint distributions would not have been changed, the mutual information should have stayed the same. That is, the identity unmixing matrix should have been the ICA solution in all cases. Hence, for all of the sources depicted in Fig. 4.1 but the first, ICA fails in separation.
On the other hand, applying ICA to sources with joint distribution exemplified in Fig. 4.2, i.e. the joint distributions belonging to the estimates of the last experiment, will yield the identity unmixing matrix in every case. However, the latter three joint distributions in Fig. 4.2 show considerable dependency levels. Yet, ICA is successful also for these dependent signals. This is obvious, considering that the sources are estimates of ICA; no more dependencies can be reduced using ICA again. The joint distributions remain the same and ICA succeeds in separation - even for dependent sources.
Fig. 4.3 summarizes these important findings in two plots. On the x-axis the dependency level is given. On the y-axis the degree of mixing is given. The angle θ adjusts a rotational mixing matrix from 0 to 45 degree. The unmixing error along the z-axis is the Amari performance index (refer to Sec. 3.3.5, Eq. 3.45). Zero equals no error and one equals maximal error in unmixing the source with corresponding joint distributions and corresponding degree of mixing.
In the left graph, utilized joints are similar to the latter 4 joints of Fig. 4.2. In the right graph, utilized joints are similar to the latter 4 joints of Fig. 4.1. Both start with little dependence and end at a high level. Note, however, that in the left graph, ICA does not act on the joints. In contrast, in the right plot, ICA act on these joints - even before mixing the corresponding signals. Specifically, Fig. 4.3 makes explicit the findings of Fig. 4.1 and Fig. 4.2. The right graph in Fig. 4.3 suggests that ICA cannot unmix the signals whatever the degree of mixing and whatever joint distributions underlie. ICA fails completely.
However, the left graph in Fig. 4.3 shows that ICA always succeeds for all degrees of mixing and all the joints underlying. Notably, even for joints with corresponding signals that are dependent. It follows from this experiment that ICA *can* separate dependent signals, if the original sources, i.e. the sources before being mixed, are dependent but cannot be made more independent by ICA.

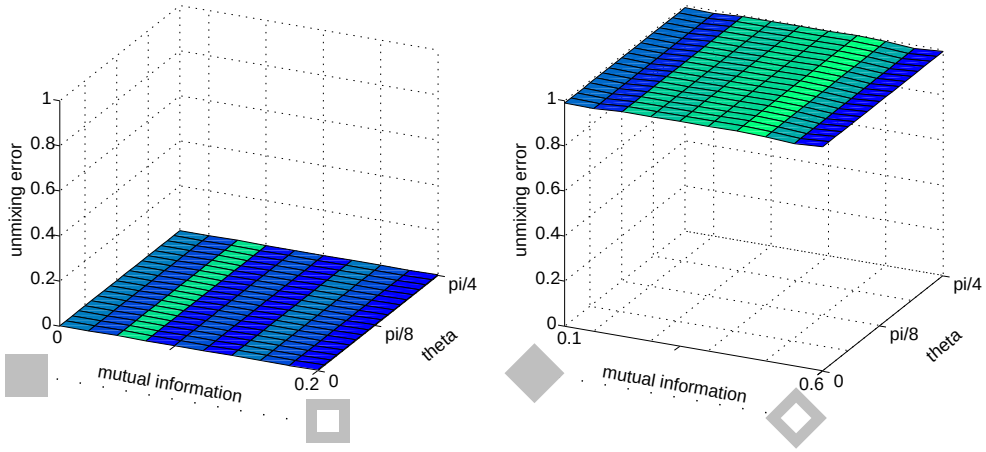


Figure 4.3: The two graphs depict two different classes of joint distributions. On the x-axis their dependency levels are given, respectively. On the y-axis the degree of mixing is given by adjusting a rotational mixing matrix from 0 to 45 degree.
The result is that ICA *can* separate dependent signals if these cannot be made less dependent by an ICA before mixing. Indeed, the source signals in the left graph have up to 0.2 dependency level but the unmixing error is zero for all mixing scenarios. Notably for no mixing, the joints cannot be reduced in dependency. Hence, ICA reduces only dependencies imposed by linear mixing and succeeds in all cases. In the right graph, ICA is observed to fail in separation even for joints with little dependencies. Notably for no mixing, dependency of all joints is reduced as the unmixing error is one. Hence, ICA reduces dependencies of the original sources and it follows that ICA fails for all mixing scenarios.

It is the bilinear model that restricts ICA from removing more dependencies. Specifically, if dependent signals cannot be made more independent by linear combinations, then linear mixing will introduce dependencies. These can be reduced by linear unmixing. Thus, dependent signals of this class *can* be recovered by ICA!
For sensory stimulation MEG data the question arises, if dependent signals exist that can be made less dependent by linear combinations before being mixed. If this is not the case, ICA will succeed even in the case of dependent sources.
Consequently, the nature of evoked dependent signals has to be assessed and the performance of ICA has to be considered carefully. In the next chapter, a virtual and an actual MEG data experiment shall be designed and extensively tested, in order to assess the performance of ICA for evoked signals.

4.1.4 The whitening step

ICA is often processed via a two step procedure. The first step is said to do 'half the job', as it cancels all dependencies based on the second order moment. Whitening has several advantages. First, it comes along with assuming independence that the variables must be decorrelated. Sec-

ond, the search of the unmixing matrix reduces to the search of orthogonal matrices. Hence, the number of degrees of freedom are reduced and algorithms become faster and more stable.
All established methods that are used in this work use whitening. Hence, whitening in the context of dependencies shall be discussed in the following.
For convenience, let me restate some equations given earlier. Specifically, whitening yielded the equation

$$\mathbf{x}_w = \mathbf{V}\mathbf{x} = \mathbf{V}\mathbf{A}\mathbf{s} = \mathbf{O}\mathbf{s}, \tag{4.2}$$

where $\mathbf{V}$ is the whitening matrix.
As whitening implies the correlation matrix to be the identity matrix, i.e. $\mathrm{E}[\mathbf{x}_w\mathbf{x}_w^T] \stackrel{!}{=} \mathbf{I}$, it follows with

$$\begin{aligned} \mathrm{E}[\mathbf{x}_w\mathbf{x}_w^T] &= \mathbf{O}\mathrm{E}[\mathbf{s}\mathbf{s}^T]\mathbf{O}^T, & (4.3)\\ &= \mathbf{O}\mathbf{C}_s\mathbf{O}^T, & (4.4) \end{aligned}$$

that

$$\mathbf{C}_s = \mathbf{I} \Rightarrow \mathbf{O}\mathbf{O}^T = \mathbf{I}, \tag{4.5}$$

must hold true and $\mathbf{O}$ is proven to be orthogonal.
However, this holds true only if the signals are independent. Otherwise Eq. 4.4 and Eq. 4.5 become

$$\mathbf{O}\mathbf{C}_s\mathbf{O}^T = \mathbf{I} \Rightarrow (\mathbf{O}^T\mathbf{O})^{-1} = \mathbf{C}_s, \tag{4.6}$$

where $\mathbf{C}_s$ is no longer the identity matrix, but can be a full regular matrix. Hence $\mathbf{O} = \mathbf{V}\mathbf{A}$ is no longer proven to be orthogonal but can be any regular matrix. Consequently, assuming $\mathbf{O}$ to be orthogonal, hence, introduces errors in recovering the mixing matrix $\mathbf{A}$, when the source signals are dependent.
However, note that in practical ICA tasks with independent signals, assuming zero cross-correlation $\mathbf{C}_s = \mathbf{I}$ is violated to some extent. The performance loss due to whitening only will not be investigated further in this thesis. Instead, the whole performance loss of ICA - including the effects of whitening - will be in focus.

4.2 Virtual MEG data experiments

In the theory section 4.1.1, it was demonstrated that ICA must have degraded performance, due to a mistake in assuming that the signals have zero cross-correlation, and due to a mistake in assuming that the signals cannot be further reduced in their mutual dependence under linear transformation. The theory section also gave examples of dependent signals that ICA can separate.

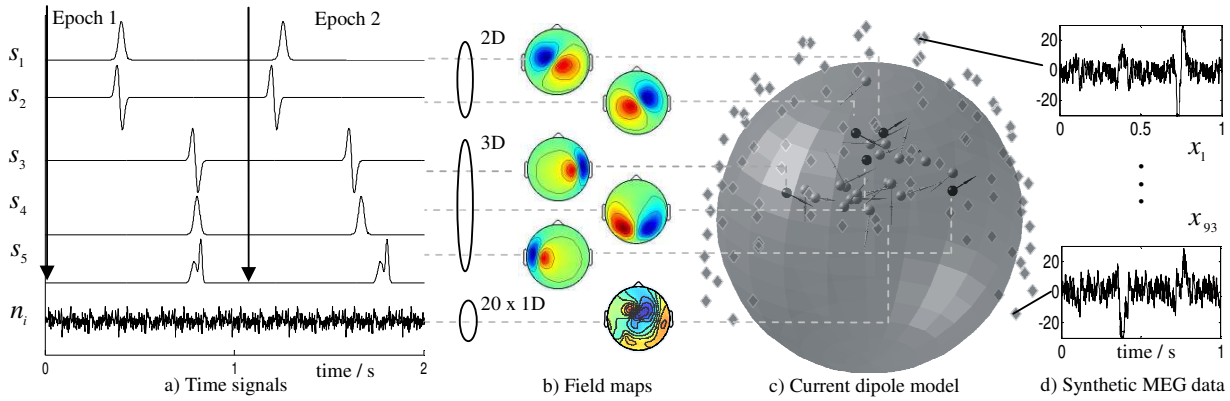


Figure 4.4: Synthetic MEG data. (a) 2 epochs of weakly dependent source signals are generated by using the SSEJR model with a high jitter value. Lowering the jitter, all epochs will eventually have responses as depicted in the first epoch leading to two independent signal subspaces (2D and 3D) with highly dependent source signals. The last row depicts one representative AR noise process. (b) Associated field maps. (c) Sphere model of human head with source (black) and noise (gray) dipoles. Sensor coordinates (diamonds) correspond to PTB 93-channel MEG. (d) 2 typical sensor observations exemplify the generated MEG data.

This section shall discuss ICA in the context of MEG sensory stimulation experiments. A virtual experiment with close to reality signals that cover a wide range of variability shall investigate the influence of dependence in this context. Subsequently, actual MEG data experiments will be considered in order to assess the performance of ICA in the actual domain when the evoked signals are tailored to be dependent.

4.2.1 The SSEJR model

In order to evaluate the performance of ICA for dependent evoked signals, it is most convenient to have a single parameter that controls the degree of dependence. The 'Synthetic Stimulus Evoked Jittered Response' (SSEJR) model [Koh09] was established for generating synthetic dependent evoked signals. The core functionality of this model is a simple mechanism that gradually changes the degree of dependency, while subspaces remain mutually independent. Against a sweep through values of the dependence parameter, the reliability of algorithms using independence (ICA), relaxed independence or no independence assumptions may be evaluated.

4.2.2 Generation of dependence

The generation process is conveniently described for two signals. Let two time signals consist of 100 trials each. Each epoch starts with an assumed stimulus followed by a time response, one for each source, that has a source specific form and latency. The response time duration is

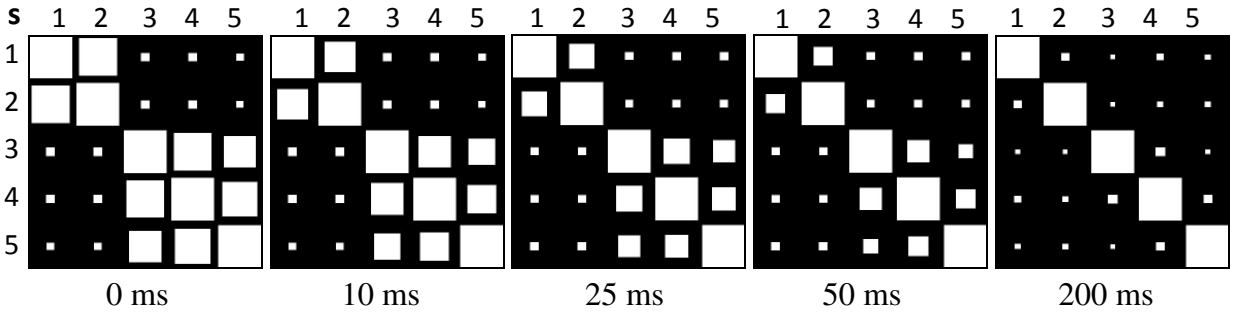


Figure 4.5: Hinton diagrams with estimates of mutual information between 5 SSEJR source signals at different jitter levels. The 2D+3D subspace setting is clearly visible. The signal dependencies decrease with increasing jitter. At 200 ms jitter all source signals can be assumed independent.

about 200 ms. A random stimulus is modeled by randomly drawing the epoch length between 800 ms to 1200 ms, such that the average epoch length is 1000 ms. As an example, one source may have a Gaussian shaped response with 300 ms latency while the other source may have a time differentiated Gaussian shaped response with 290 ms latency. This setting is depicted in the first epoch of Fig. 4.4 a). When all epochs are equal to epoch 1, the resulting signals in the 2D subspace are highly dependent. In order to lower the dependencies, the source specific latencies are changed in each epoch by two normally distributed random numbers with zero mean and standard deviation σ, which leads to a jitter in the latencies. In the following, σ is referred to as the jitter parameter. It is the key parameter of the model to gradually control the degree of dependence. The greater the jitter, the lower the dependency between the signals and vice versa. This model was named the Synthetic Stimulus Evoked Jittered Response (SSEJR) model. Any subspace of any size may be modeled likewise. Subspaces are made mutually independent by choosing all source specific latencies such that signals from different subspaces do not overlap in time. In this section, a setting uses 5 signals in 2D and 3D subspaces are used and depicted in Fig. 4.4. The dependence of a set of SSEJR generated signals s_i, $i = 1, ..., 5$, will be evaluated in terms of their normalized pairwise mutual information, which may be expressed as

$$\mathrm{I_n}(s_i, s_j) = \frac{\mathrm{I}(s_i, s_j)}{\sqrt{\mathrm{H}(s_i)\mathrm{H}(s_j)}}, \tag{4.7}$$

where $\mathrm{H}(s_i)$ is the entropy of s_i. Mutual information and entropy were calculated based on discrete bin-wise evaluated marginal and joint distributions [Kra04]. Note that the marginal distributions $\mathrm{p}(s_i)$ and $\mathrm{p}(s_j)$ of any pair of signals within a subspace remain unchanged with varied jitter. The joint distribution $\mathrm{p}(s_i, s_j)$ approaches the product of the marginal distributions while increasing the jitter as more sample combinations occur. Hence the mutual information decreases with increasing jitter. Fig. 4.5 shows the normalized pairwise mutual information between 5 simulated SSEJR source signals for different jitter values. Particularly, the Hinton diagrams show a block diagonal structure with small off-blockdiagonal elements indicating almost mutually independent subspaces. Elements associated with source signals within each subspace decrease

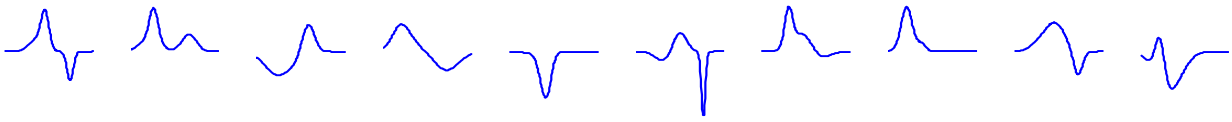


Figure 4.6: 10 examples of randomly generated evoked responses using a mixture of 3 bell shaped time signals. Notably, a large variety of simulated evoked signals can be generated at random allowing to test a whole class of various smooth evoked signal forms against separation performance of the method under test.

gradually with increasing jitter, which confirms the expected behavior of the SSEJR model.

4.2.3 Signal generation

Fixed signal forms as shown in Fig. 4.4a may not be general enough. As a matter of fact, the use of different signal forms influences the performance of methods under test. Ideally, one would like to consider all signal forms that can occur. The time domain signals are generated as follows. In the e-th iteration, for a fixed jitter level, the f-th source signal shape is given by

$$s_f^e(t) = \sum_{k=1}^{3} a_k \ \exp(-\frac{(t-m_k)^2}{s_k^2}), \tag{4.8}$$

where t is the time index with a support of 200 ms, and a_k, m_k and s_k are uniformly distributed random numbers with $-1 < a_k < 1$, $-50 < m_k < 50$ and $5 < s_k < 50$, respectively. Using Eq. 4.8, a large variety of simulated evoked signal shapes can be generated at random. A draw of 10 responses is depicted in Fig. 4.6.

4.2.4 Setup and data assessment

A homogeneous conducting sphere model is taken for an approximation of the human head and the neuronal currents in the brain are modeled by equivalent current dipoles (ECD) [Sar87, Häm93]. As an example, 25 dipoles may be modeled, 5 dipoles to represent evoked sources, and 20 dipoles to represent independent noise processes. The latter ECDs shall be placed at random, whereas the signal dipoles may be placed as to represent physiologically meaningful positions. Each noise dipole represents a source that follows a 7-th order AR process activated by Laplacian distributed white noise with AR coefficients that were obtained from real world prestimulus MEG data. Synthetic MEG data are obtained by a processing chain depicted in Fig. 4.4. The ECDs are placed in the spherical human head model. Magnetic fields are calculated using PTB 93-channel MEG sensor coordinates, together with quasi-stationary Maxwell equations for dipole currents [Sar87]. Following the linear ICA model $\mathbf{X} = \mathbf{AX}$ of Eq.3.3, the mixing matrix $\mathbf{A}$ emerges from the contribution of each source to each sensor. The superposition of all magnetic fields gives the observed MEG data $\mathbf{X}$. Finally, sensor noise is introduced with an average SNR of 30dB. Fig. 4.4b depicts associated field maps and Fig. 4.4d depicts typical sensor observations generated.

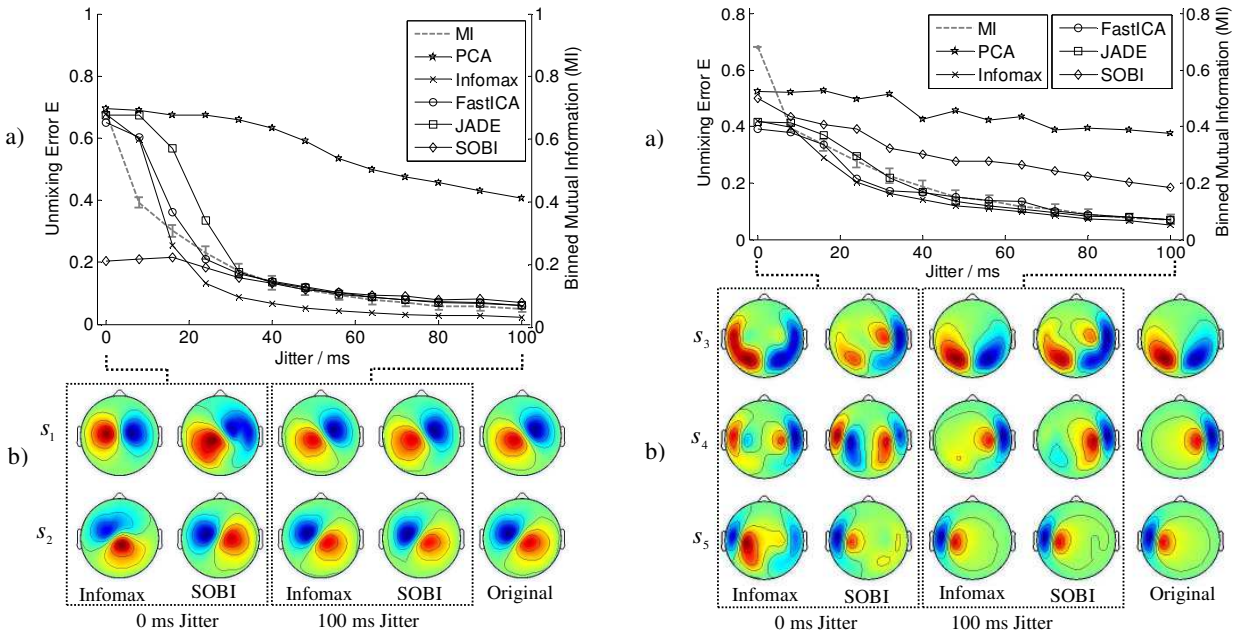


Figure 4.7: Performance of ICA (a) and associated recovered field maps (b) for the 2D subspace (left) and 3D subspace (right) using signal forms and source constellations as depicted in Fig. 4.4. All methods recover almost independent sources for high jitter values; the associated field maps are close to the originals. However, it can be observed that all methods lose performance when the jitter value is lowered; the unmixing error closely follows mutual information. The figures show the performance of one specific setting, with random jitter values. Results with random source locations and random signal forms are given in the comparison section at the end of this work (Figs. 5.15 to 5.18).

4.2.5 Results

Let the performance of ICA be analyzed for a first source setting, that of Fig. 4.4, i.e. with two independent subspaces containing two and three dependent signals, respectively. These results represent ICA performance for a specific setting, as the signal forms do not vary. The virtual experiment is repeated 50 times per jitter value, yielding 50 realizations for each jitter value. The signal locations and signals forms are the same for all experiments. Performance is evaluated using median unmixing error and some corresponding field maps.

Particularly, two sets of SSEJR signals are analyzed by running 50 realizations per jitter value. The median unmixing error and the median recovered field maps are depicted in Fig. 4.7. The left graph shows the ICA performance of unmixing the 2D subspace, while the right graph suggests the ICA performance of unmixing the 3D subspace. In the left graph, the original sources have a bell shaped and a time differentiated bell shaped response, whereas in the right graph the original sources have a bell shaped and a temporal differentiated bell shaped response as well as a mixture of 3 bell shaped time signals with fixed locations (cf. Fig. 4.4a and 4.4c).

Normalized mutual information, Eq. 4.7, are calculated from the generated signals. Fig. 4.7 sug-

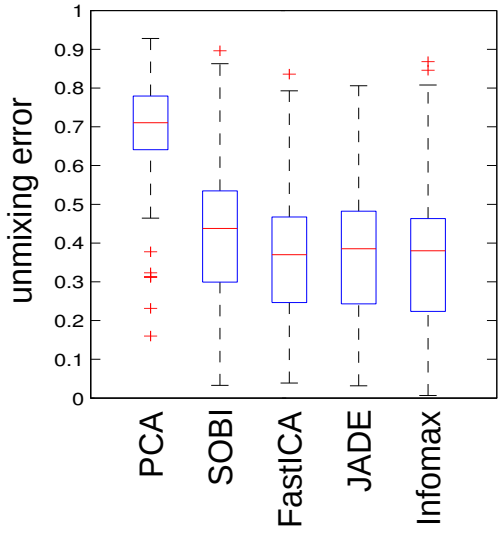


Figure 4.8: Random signal shapes generated according to Eq. 4.8 and dependent signals (low jitter values) show that ICA fails on average. Some signal configurations can be separated by ICA. However, on average, the performance is not good enough asking for alternatives to separate evoked dependent signals.

gests for both graphs that the median normalized mutual information decreases with an increasing jitter value in line with our demands on the SSEJR model. For an assessment on separation performance the unmixing error and typical field maps are provided. Below each graph, the original source field maps are depicted. As expected, the performance of principle component analysis (PCA) is poor even in the vicinity of independent source signals, i.e. for high jitter value. This supports the well known fact that decorrelating generally does not lead to a correct separation. In contrast, all established ICA methods show a good performance in recovering almost independent sources. In both figures the unmixing error drops under 0.2 for 100 ms jitter; the associated field maps are close to the originals.
However, it can be observed that all higher order statistic based ICA methods lose performance when the jitter value is lowered. Strikingly, their unmixing error closely follows the binned mutual information. The corresponding estimated field maps obtained by the Infomax methods is depicted for 0 ms. These maps suggest that the source field maps have not been found correctly. Most clearly visible, in the right graph, all Infomax recovered patterns show contributions from all source field maps - the separation task failed.
In contrast to the distribution processing ICA methods, the performance of SOBI is more sensitive to the shapes of the signals. From the left graph in Fig. 4.7, it can be observed that SOBI performs best for the signals simulated in the 2D signal subspace. However, from the right graph in Fig. 4.7, SOBI is observed to perform worst. The finding that SOBI does not closely follow the normalized mutual information may be explained as follows. SOBI exploits independence of the lagged time signals. Yet the SSEJR model systematically adjusts the mutual information only for 0 lag, while the mutual information of lagged version of the generated signals may vary. Hence, it depends on the choice of time lags, how SOBI performs. However, even for wisely chosen time

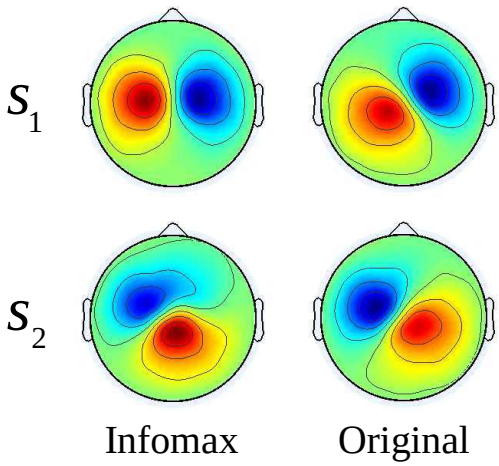


Figure 4.9: Example of closely spaced dipoles: the recovered field maps for dependent SSEJR sources do not correspond to the original sources. Hence, physiologically meaningful estimates do not prove the success of ICA and have to be interpreted carefully.

lags, which is a difficult task in the actual data scenario, SOBI has problems with similar power spectra, which occur when no jitter is present but which are less for highly jittered responses.
In summary, Fig. 4.7 suggests that all tested ICA methods show a decreasing performance for increasing dependencies between two source signals. The results cannot suggest a preference for a specific ICA method as performance is highly influenced by the signal shape.
Repeating the experiment for low jitter and randomly chosen signals yields the results depicted in Fig. 4.8 (see also Sec. 5.7) The figure suggests that all methods succeed for some shape configurations. Nevertheless, ICA fails on average, i.e. for most signal configurations tested.

4.2.6 The near dipole effect

Yet another aspect is revealed by changing the sites of the source dipoles. For closely spaced dipoles, all established ICA methods recovered clear dipolar patterns, which were different from the originals. As an example, in Fig. 4.9 the recovered pattern from dependent sources, significantly differs from the original source field map. This pattern seems to emerge from a single dipole but in fact is still a linear mixture of the source field maps. Hence, the common practice of judging an ICA decomposition to be successful when the estimated source patterns suggest physiologically meaningful results may be wrong. This is an important result, exemplifying that the plausibility of the structure of an ICA estimate can be misleading. For closely spaced activities, mixtures can show a dipolar pattern, as suggested by Fig. 4.9. Hence, physiologically meaningful ICA estimates must not be confused with successful source separation. Interpretation of the estimates has to be done with care, especially considering that the results represent *some* vector of a subspace of evoked signals.

4.3 Actual MEG data experiments

Virtual experiments are a vital tool in gaining understanding on different features of data and to assess the separation method due to a well-defined ground truth. However, signal processing methods also have to be tested in real world scenarios. For our project, it was of importance to design a suitable paradigm, allowing to assess ICA in a dependent MEG data scenario. From the dependency theory section, it can be concluded that this is best accomplished when signals have temporal overlap.

4.3.1 An audio-visual paradigm

Inspired by the work of Tang et al. [Tan05] a novel actual data experiment was designed. In Tang's work, median nerve stimulation was used to test the source localization performance of SOBI. Left, right and bilateral median nerve stimulation were recorded. The localization result from bilateral stimulation was compared to the unilateral stimulation that was taken as a reference. The utilized N20 signals that occurs after median nerve stimulation is known to be stable. However, bilateral evoked N20 signals only have short temporal overlap and may not be very dependent if small time shifts occur.
For our purpose, a paradigm was designed using auditory and visual stimulation in order to obtain significant temporal overlap. Nevertheless, in line with [Tan05], unisensory auditory stimulation is used as reference and so is unisensory visual stimulation. Using noise adjusted PCA, a reference mixing vector for only auditory stimulated data and a reference mixing vector for only visual stimulated data is obtained. Against these reference mixing vectors, the separation results obtained from bisensory audio-visual stimulation data is compared. Assuming the location of unisensory evoked activity and bisensory activity the same, the Amari unmixing error can be used, as the reference mixing vector are taken as the original mixing matrix. Note that in contrast to the locations, the temporal profiles of the sources are allowed to vary. This is important, as influence from visual stimulation to auditory processing has been discussed [McG76].

4.3.2 Subjects and recordings

Ten right-handed subjects from an homogeneous group of 5 male and 5 female students aged 20 to 30 participated, voluntarily. None of the subjects reported neurological pathologies. All were healthy at date of recording. All subjects were informed about the recording facility and non-invasive procedure. During recording a communication system was running, such that information could be given into the cabin and received from the subject. All subjects were instructed to give a sign if not feeling well. One experiment had to be cancelled, due to magnetic material that could not be removed from the subject. In total, 9 datasets were recorded, each with a total duration of 20 minutes.

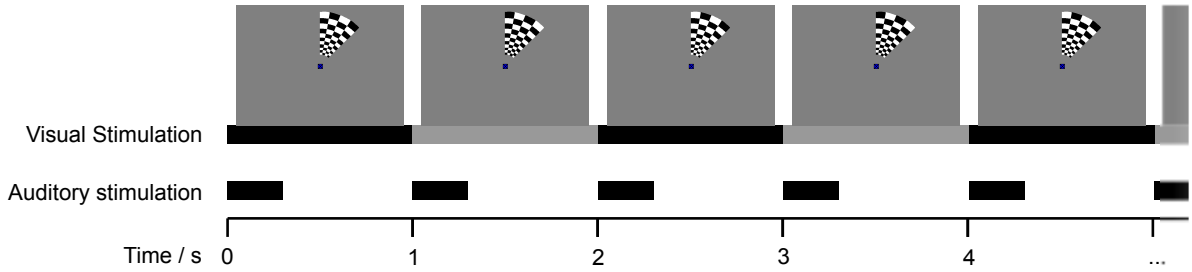


Figure 4.10: Graphic representation of audio-visual stimulation paradigm. Tones of 1 kHz with a duration of 300 ms were presented as well as alternating checkerboard wedges. Trial length varies in reality between 1 to 2 seconds. Visual (V) and auditory (A) and audio-visual (AV) stimulation are presented to the subject in an interleaved and randomized fashion.

4.3.3 Experimental setup

Evoked MEG data with simultaneous auditory and visual stimulation are recorded using a full head 125 channel MEG system [Kad99]. Presentation® (Neurobehavioral Systems, Inc.) is used with the following settings.
Tones of 1 KHz with a duration of 300 ms [Häm93, Har05, Nää87, Sal04] were presented monaurally to the right ear of subjects. Checkerboard wedges [Slo99, Ahl92] were presented in an alternating fashion [Ain03, Sha00] to the right visual field. By this, stimulation was optimized to mainly evoke auditory or visual related activity in the left cerebral hemisphere [Oka82].
The duration of the single-trials lasted between 1 to 2 seconds [Slo99, Ain03]. Visual (V) and auditory (A) and audio-visual (AV) stimulation was presented to the subject in an interleaved and randomized fashion. The sampling frequency was set to 500 Hz. The software Presentation® recorded trigger levels, from which the single-trials could be identified. Specifically, 300 trials were recorded and subsequently cut from the continuous recordings. This data was bandpass filtered with an zero-phase filter [Pic95] between 0.5 Hz and 100 Hz. The concatenation of single-trials with 100 ms prestimulation duration and 300 ms post-stimulation duration yield the MEG data utilized. All trials were detrendet [Wüb07] and the offset was removed [Pic95]. Bad trials were identified after an ICA [Del04] and removed, subsequently. This was done in a fashion that the best 200 trials were taken, each having 400 ms length. The paradigm is outlined in Fig. 4.10; the full audio-visual paradigm code using Presentation® is provided in the appendix.

In Fig. 4.13, trial-average per channel of the unisensory and bisensory evoked data (A,V and AV) is plotted for all 9 datasets. Normalization was done by dividing trial-averages by the largest absolute value of temporal instants and over channels. Corresponding field maps are depicted for one time instant in order to visualize the spatial field distribution. Fig. 4.13 suggests different levels of signal quality. While the first data suggest high SNR, the last data has a clearly lower ratio. Nonetheless, all unisensory evoked field maps show typical patterns. The auditory evoked fields show dipolar structures over the primary auditory cortex, contra- and ipsilaterally. The

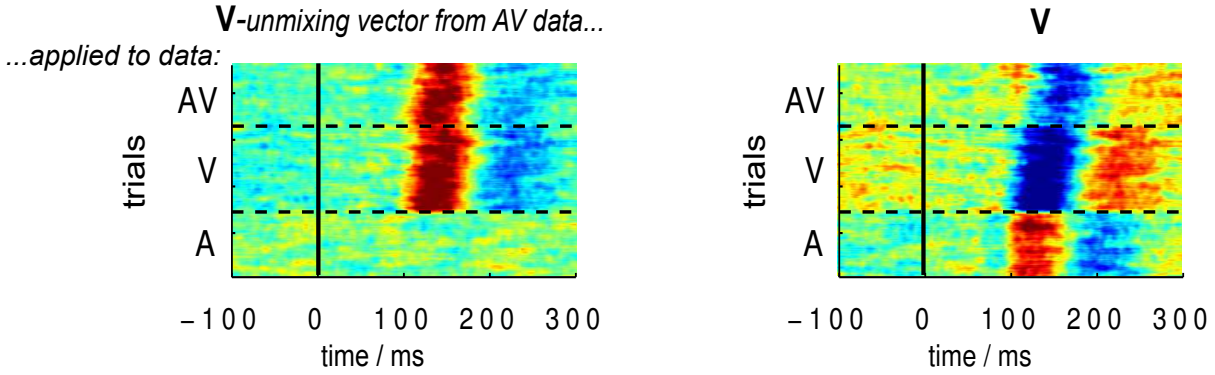


Figure 4.11: A composite single-trial plot is obtained by applying one unmixing vector - found using AV data - to AV data and to A sand V data, respectively. If, e.g., the V unmixing vector separates the visual component from AV data, then similar evoked activity should be visible when applying the V unmixing vector to AV data as well as to V data. In contrast, applying the V unmixing vector to A data, no evoked contribution must be observable (left figure). If an auditory evoked activity is observable, this auditory activity is mixed into the component found by applying the V unmixing vector to the AV data. Hence, unmixing is not perfect, which is observable by this composite single-trial illustration technique (right figure).

visual evoked pattern show dipolar structures over primary visual cortex. As intended, a clear tendency to the left hemisphere is observable.

From the patterns of the bisensory audio-visual MEG data, the experimenter cannot clearly determine the visual and the auditory fields. Some show the mixed distribution, others only give evidence to one of the two stimulation modalities, namely auditory and visual. Blind separation shall be applied to recover the underlying original source signals and the corresponding unmixed filed maps.

For the sake of source separation assessment and development in the actual domain, uni- and bisensory stimulated data were preprocessed further. The unisensory stimulated data is projecting on a single mixing vector, obtained from comparing post-stimulation and prestimulation energy. Noise adjusted PCA (NAPCA) (refer to Sec. 5.1.2) is used to obtain two one-dimensional reference vectors. The obtained mixing vectors, i.e. one from unisensory auditory evoked data (A) and one from unisensory visual evoked data (V), give the reference mixing vectors for the bisensory evoked data. Specifically, the bisensory audio-visual stimulated data is projected onto the two-dimensional subspace spanned by these reference vectors. Hence, the bisensory evoked data (AV data) is two-dimensional after this preprocessing. The two-dimensional mixing matrix is known to be the concatenation of the auditory and visual reference vectors and the Amari unmixing vector can be used, assuming that the location of unisensory and bisensory evoked data are the same.

4.3.4 Novel performance illustration technique

The actual data of nine subjects form the basis of evaluating the algorithms in the real world scenario. As a consequence, every source separation method can only act on a small number of data sets. The unmixing error err is a functional of a matrix, mapping quality to a single number. An unmixing error of say 0.3, does say something about the average separation performance, but it may not be descriptive enough for a single recording.
Hence, the quality of actual data separation shall be further illustrated by comparing single-trials from A,V and AV data for more detailed performance analysis. For this, recovered single-trials are stacked row-wise in a matrix. This matrix in turn is displayed in a color coded fashion as in Fig. 4.11 [Jun01]. Let the unmixing vectors of a separation algorithm be labeled as V if the visual single-trials are recovered and as A if auditory single-trials are recovered by applying the unmixing vector to the mixed data. The A and V unmixing vectors are obtained by running the separation method under test using only AV data. Separation performance can be analyzed by stacking single-trials recovered from the AV data as well as single-trials recovered from the A data and the V data by applying each time the same unmixing vector recovered from using only the AV data.
If the V unmixing vector obtained from AV data separates the visual component, then an evoked activity should be obtained applying this V unmixing vector to the AV data as well as to the V data. In contrast, no evoked activity should be obtained, when applying the V unmixing vector to the A data. If evoked activity is observable from applying the V unmixing vector to the A data, unmixing is not perfect.
In Fig. 4.11 this novel illustration technique is displayed. Clearly, the left graphic shows a successfully recovered visual component. The V unmixing vector gives an evoked contribution when being applied to the AV data; applying it to the V data yields a component with similar single-trial structure. Applying the V unmixing vector to the A data yields no evoked contribution. Hence, a very good separation is observable.
In contrast, the right graph shows a separation that has not been successful. The V unmixing vector applied to A or V data yields a contribution each time. In fact the contribution from A and V data are of opposite signs and, hence, applying the unmixing vector to the AV data yields a mixed component with auditory and visual contributions that almost cancel.

4.3.5 Results

In Fig. 4.12, the boxplots suggest that ICA does not sufficiently separate the audio-visual data. All established methods have high median unmixing error. Hence, separation did not succeed on average. Note that this is in contrast to the findings in the work of [Tan05].
Furthermore, using the novel performance illustration technique, Fig. 4.14 holds the results using FastICA. Clearly, the novel illustration technique gives insight into how the different datasets have been separated.
The optimal estimates are obtained by using the reference mixing vectors concatenated to a pseudo mixing. Inversion gives an unmixing matrix. Using the vectors from this unmixing ma-

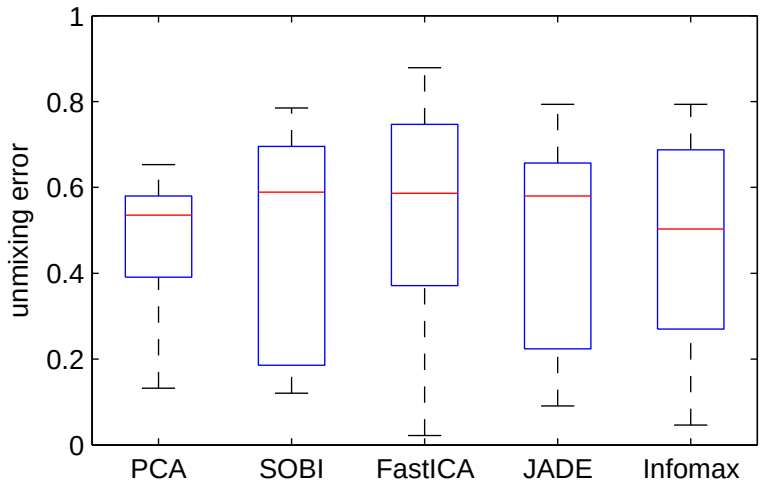


Figure 4.12: The boxplots suggest that ICA does not sufficiently separate the AV data. All established methods have a median unmixing error err, which indicates that separation did not succeed.

trix yields by construction no contribution from A when the V unmixing vector is used and vice versa. However, it is worth comparing the evoked activity using AV data to using only V data or only A data. For example, for subject 8, a very similar temporal profile of the single-trials are found from the AV data and the unisensory evoked data. For the visual component, analyzing subject 1, the activity seems to be more jittered when obtained from the bisensory evoked data. All data showed reasonable agreement between single-trials from unisensory and single-trials obtained from bisensory evoked data. Data from subject 7, observing the visual component, seems to have the most disagreement.
The ICA estimates show a completely different picture, suggesting the usefulness of the novel performance illustration technique. For example, the stacked single-trial pictures obtained for subject 9 using ICA, show evoked activity for both the auditory as well as the visual unmixing vector. However, a comparison with the unisensory evoked data shows that both recovered activities are still mixtures of auditory and visual signals. This can be seen as the auditory unmixing vector should not yield activity when being applied to the visual evoked data. Hence, unmixing failed.
In summary, the recordings from subject 6 are perfectly recovered using ICA; the recording from subject 7 show a good performance. However, the remaining recordings fail in separation for one or even both components. Fig. 4.12 suggests that actual evoked MEG data can have dependent evoked signals, and that ICA is able to lower the contrast of independence for these signals under linear transformations. This, however, leads to a failure in ICA decomposition as it was demonstrated in Sec. 4.1.3.

4.4 Discussion

Statistical dependence is difficult to measure; full knowledge of marginal and joint probability distributions are involved. Indeed, numerous methods aim at estimating independence in a robust and efficient manner [Kra04]. Each of these methods can be used to set up a contrast for ICA yielding numerous ICA methods. Yet, the most important *feature* of statistical independence is its strictly non-negative property and that it must be zero only if the joint distribution factorizes into the marginal distributions. Hence, an ICA contrast must obey this feature. All established ICA methods do have in common to optimize independence. In order to drive the components towards independence, they use different features of mutual information that differently act on the data.

It was demonstrated that ICA *can* separate dependent source signals. This is possible, in the case that the source signals are not changed by ICA if these are directly fed to ICA before being mixed. Thus, ICA cannot further optimize its contrast function for these dependent source signals. On the other hand, an ICA methods fails if its contrast acts on the original signals in a way that it mixes the sources to optimize the contrast. An optimal contrast is invariant to all source dependencies and only reduces dependencies introduced due to linear mixing.

In this chapter, all established methods yield bad median performance for dependent signals. Hence, all established methods lower their utilized contrast function even for the task of separating the sources. However, all established ICA methods possess cases that are fully successful, although the sources are dependent. In line with the comparison section, about 25 % of our signal configurations tested in the two-dimensional setting show results below 0.25 unmixing error. This means 25 % of ICA results are significantly better than random unmixing choosing a 0.1 significance level. The conclusions are manifold. First, ICA is not robust against dependencies. Second, ICA can fully separate dependent signals in some cases and yields approximate results in about one out of four results. One may conclude that ICA cannot be used for dependent signals. However, ICA should be considered as a preprocessor in order to obtain good starting points for any method that is dependent on these.

In order to further assess performance, an actual audio-visual MEG experiment was set up. A novel performance illustration technique shows that the separated single-trials using audio, visual and audio-visual data matches if the constructed 'original' mixing matrix was used. The illustration technique allows to evaluate ground truth in multisensory evoked data with unisensory evoked reference data. Indeed, a high level of matching in all 9 data sets is observable suggesting the high value of these recordings for the assessment of source separation performance. The results show that ICA is not suited to separate these actual data. Only two data sets show activities that matches for both components, i.e. that may be considered separated.

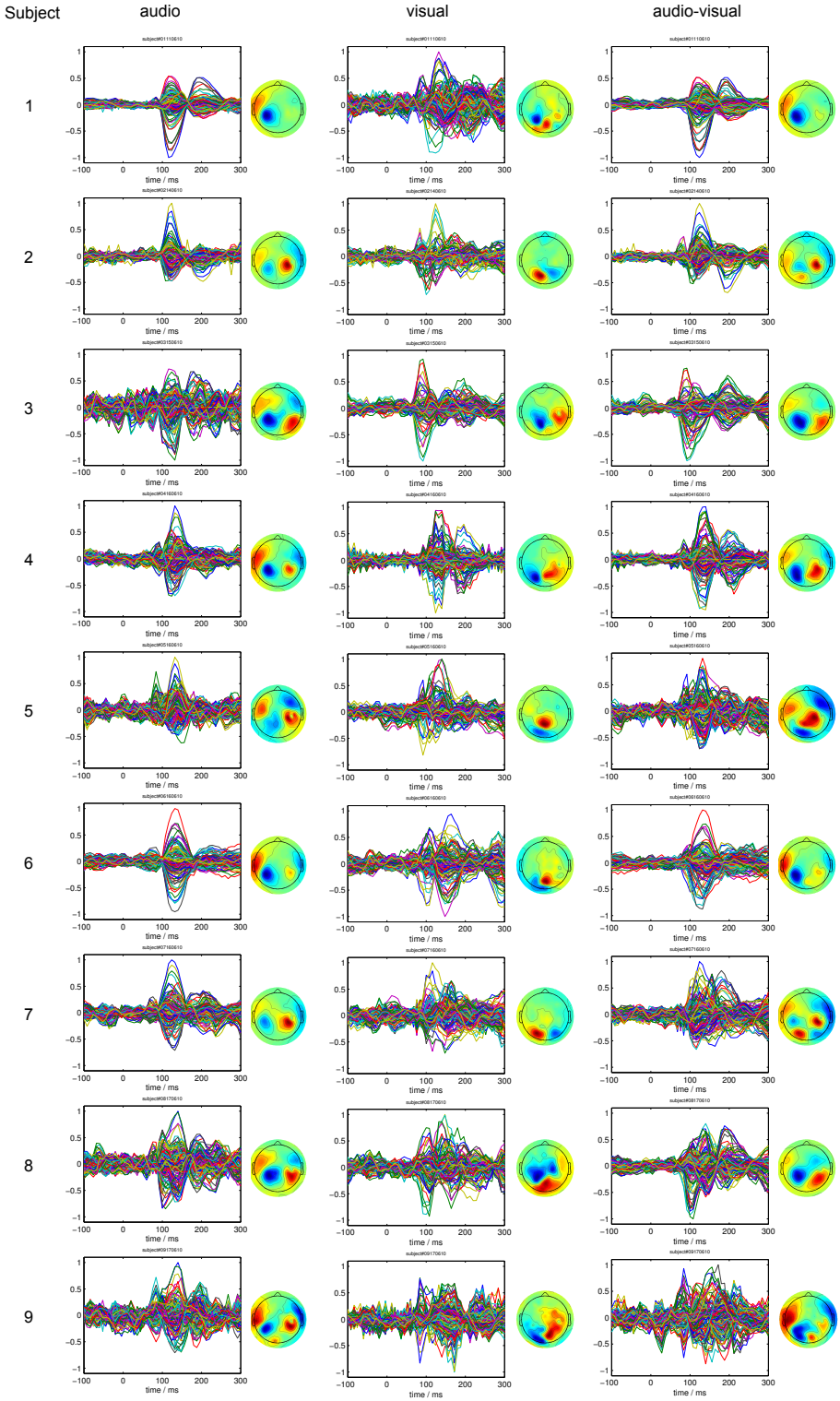


Figure 4.13: Trial-averages per channel for all subjects utilizing actual auditory evoked (A) data, visual evoked (V) data and audio-visual evoked (AV) data. In each plot, for each channel, 100 trials of 400 ms duration were averaged. One corresponding field map at maximal normalized field strength suggests the location of evoked activity. As expected, clear structured field maps are obtained for A and V data; less structured field maps are obtained for AV data.

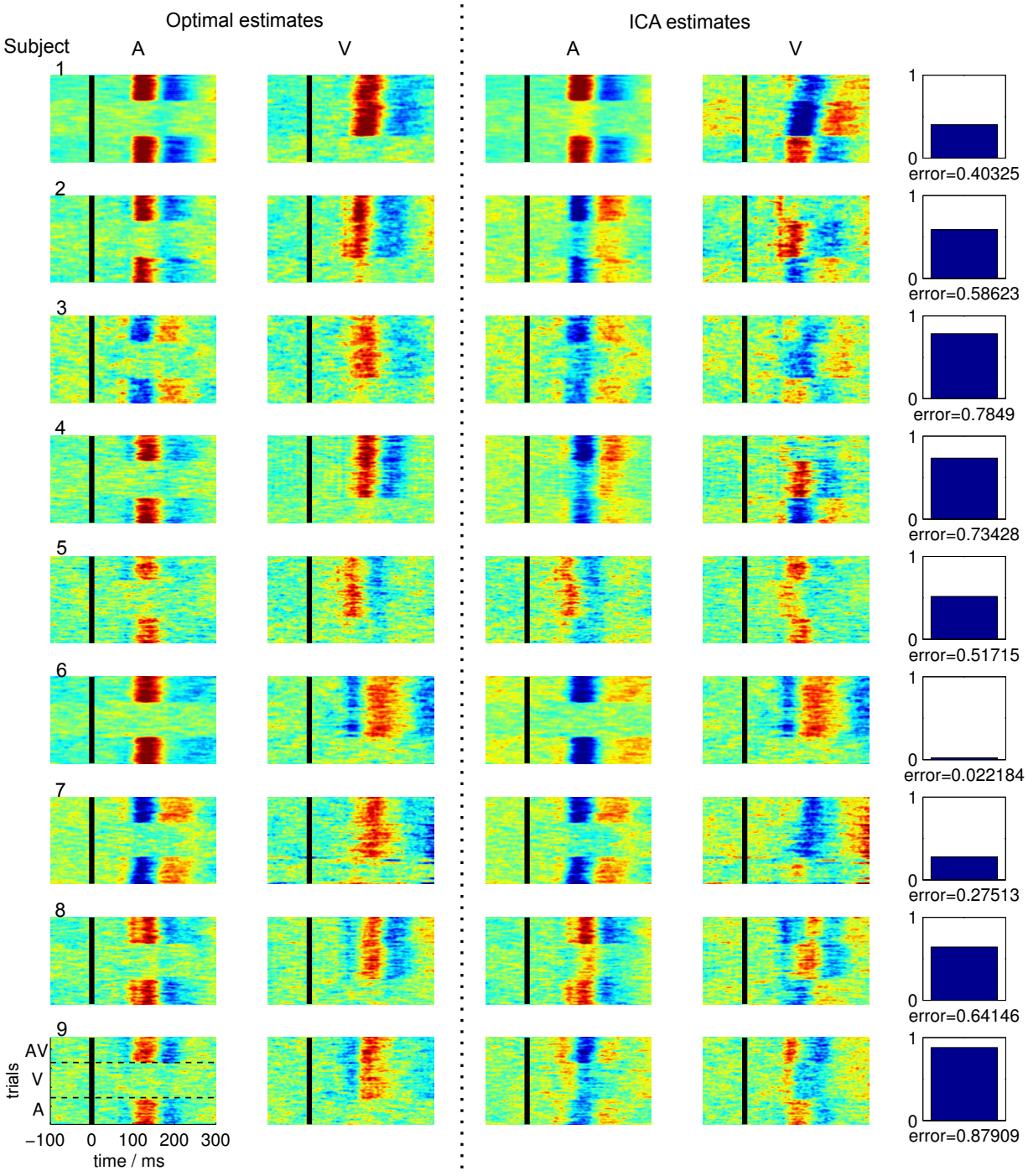


Figure 4.14: Results of applying FastICA to the data from the audio-visual experiment (AV data). The FastICA unmixing vectors are applied to all AV, A and V data, in order to assess the performance. In the rightmost column the unmixing error is also provided. The obtained single-trial plots show that only two recordings are separated by FastICA. Hence, actual domain data are shown to be dependent in a way that ICA can fail.

5 Techniques for the separation of evoked dependent source signals

Next to the established ICA methods, a large variety of decomposition algorithms exist, e.g. [Abr03, Cai06, Hyv04]. Many other algorithms tackling the separation of dependent signals are problem dependent as they use special knowledge, such as sparsity, time-, frequency-, phase- or energy-contents. Furthermore, all methods that operate under the bilinear model have to make assumptions that ensure essential uniqueness. Specifically, a method assuming a dependence contrast may be more robust to the violation of independence but will run into problems if this dependence contrast is not met.
In contrast, here, extended trilinear models shall be used. It is known that evoked MEG data are highly structured in space and time as well as in trials. Trilinear models make use of this structure yielding essentially unique decomposition without the need of assumptions such as independence [Har70, Möc88, Fie91, Mor08]. However, due to shifts, trilinearity can be violated and it is demonstrated that this approach may not work, either.
In this chapter, the recovery of the subspace of evoked dependent signals is discussed and trilinear models are considered for the task of evoked dependent source separation. Subsequently, a novel technique is presented that is not trilinear as it also models signal shifts. For this, a model is used that is more flexible than a trilinear model, while still being essentially unique. It is demonstrated that evoked data can be explained by this model and that separation of independent as well as dependent sources is possible without the need of extra assumptions, such as independence.
Finally, a comparison study shall assess the performance of all established ICA methods, generalized ICA methods, trilinear methods and our novel time domain shifted factor analysis (TDSFA) method. For assessment, virtual MEG data and actual MEG data are carefully set up and evaluated.

5.1 Evoked subspace analysis

In order to reduce the complexity of dependent source separation, an important preprocessing step is the recovery of the subspace of evoked dependent signals. After a brief definition, independent subspace analysis (ISA) and noise adjusted principal component analysis (NAPCA) are considered and compared using synthetic data.

Again, let $\mathbf{X} = \mathbf{AS} + \mathbf{E}$ be the recorded MEG data, where $\mathbf{E}$ denotes sensor noise. Let the i'th column of the mixing matrix be $\mathbf{A}_i$ and the i'th row of the signal matrix be $\mathbf{S}^i$. Let further Ω be the set of signal indices and EV the subset of evoked signal indices with $EV \subset \Omega$. The recorded

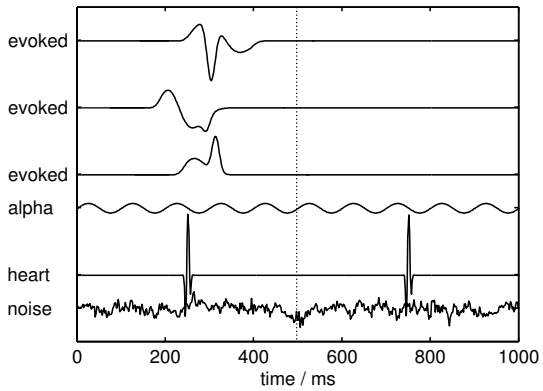


Figure 5.1: One trial of synthetic evoked MEG signals. The first three rows depict typical evoked signals, while the remaining rows depict background activities. The latter reflect possible non-evoked signals. Noise adjusted PCA (NAPCA) utilizes two covariance matrices that are estimated from pre- and poststimulation data. Based on these noise adjusted covariance matrices, the evoked subspace is found via a sequence of whitening and ordinary PCA.

MEG data can be represented by two parts, given by the evoked signal part

$$\mathbf{X}^S = \sum_{i \in EV} \mathbf{A}_i \mathbf{S}^i \tag{5.1}$$

and by the noise part

$$\mathbf{X}^N = \sum_{i \notin EV} \mathbf{A}_i \mathbf{S}^i + \mathbf{E}, \tag{5.2}$$

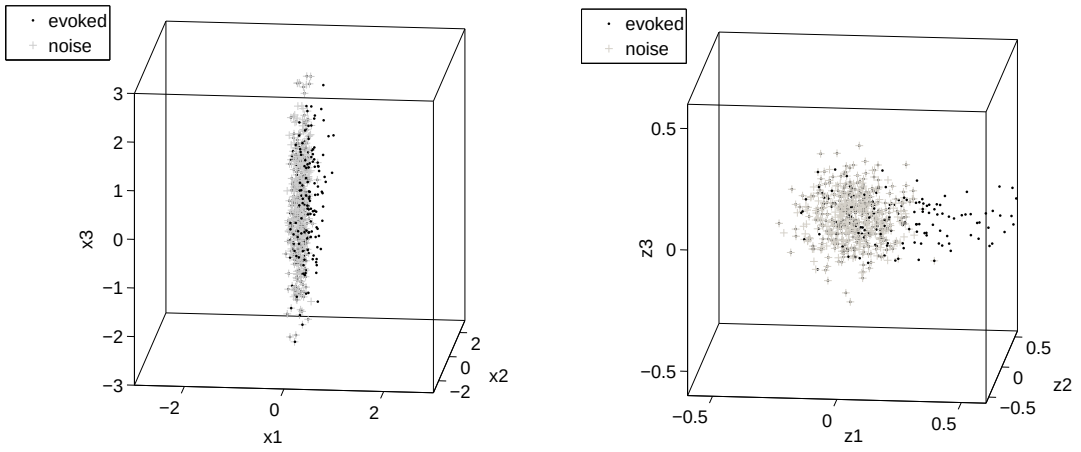
where the latter summarizes all non-evoked signal contributions. Evoked MEG data may be expressed as

$$\mathbf{X} = \mathbf{X}^S + \mathbf{X}^N. \tag{5.3}$$

The task of evoked subspace analysis is to find a basis, i.e. any set of linear independent vectors that are linear combinations of the evoked mixing vectors $\mathbf{A}_i$, and to find the corresponding linear combinations of the evoked signals $\mathbf{S}^i$, where $i \in EV$.

5.1.1 Cardoso's conjecture

Independent subspace analysis (ISA) [The06, Gut07, Hyv01b] aims at subspaces of sources that are dependent, while sources from different subspaces are assumed to be mutually independent. Hence, ISA is a generalization of ICA in that the sources are allowed to be dependent within a group as long as different groups are mutually independent. Specifically, Cardoso proposed to use



(a) (b)

Figure 5.2: Mechanism of NAPCA. Ordinary PCA cannot recover the evoked subspace directly from the data (a). By whitening the data (b) based on prestimulus data, the evoked subspace is found by PCA based on whitened poststimulus data. Due to whitening, the eigenvalues corresponding to noise are one, while the eigenvalues corresponding to evoked signal must be greater one by definition. Notably, next to finding the evoked subspace, its dimensionality can also be inferred.

ICA followed by grouping the estimates to perform ISA [Car98b, Sza06]. Thus, he conjectured that having data with underlying independent and dependent signals, all signals are separable by ICA but the dependent signals. If Cardoso's conjecture is valid, ISA can be performed in two steps. After running ICA, mutual information of the estimated sources is used to group the estimates into subspaces of dependent signals and subspaces of independent signals. In contrast, it is argued in section 5.3 that mutual information of estimates may not be easily linked to mutual information of the sources and that this technique can fail.
Other ISA methods model the dependent subspaces but either suffer from high combinatorial complexity [Poc05] or assume that the dimensions are known [Hyv06]. Some efforts have been made to overcome this drawback [Bac03, Poc06] by introducing further signal assumptions. These, however, may not hold for all evoked signal settings.
Consequently, the use of noise adjusted principal component analysis (NAPCA) is advocated for finding the evoked signal subspace. NAPCA uses simple PCA techniques while considering information from pre- and poststimulation, which is detailed in the following.

5.1.2 Noise adjusted principal component analysis

According to Equation 5.3, evoked MEG data can be split into a noise and a signal part. NAPCA [Lee90, Wüb07] assumes that the noise part is stationary, that the signal part is zero at known data periods, and that the noise and signal part are uncorrelated. This setting is depicted in Figure 5.1. Let $\mathbf{C}_{\mathbf{X}}$ and $\mathbf{C}_{\mathbf{X}^N}$ be the covariance matrix of the data and noise part, respectively. NAPCA then

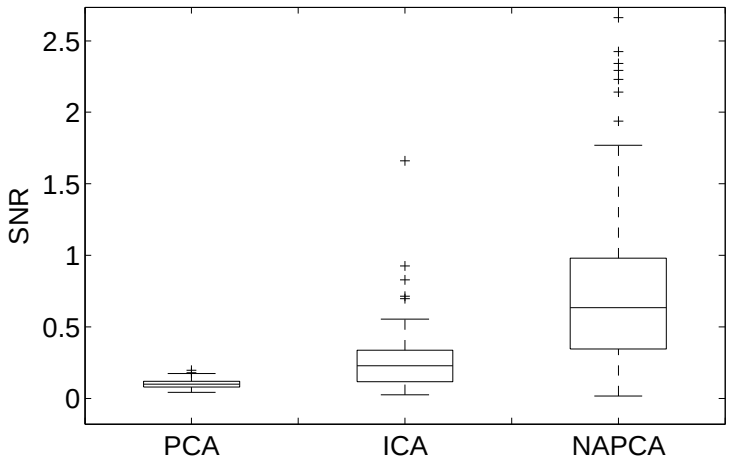


Figure 5.3: Results from virtual experiments on the SNR performance [Lem06] of recovering the subspace of dependent evoked signals using PCA, ICA followed by SNR based grouping and by NAPCA. Clearly noise adjusted NAPCA yields best performance.

consists of two linear transformations. The first transformation whitens the noise part by left-multiplication of Equation 5.3 with a whitening matrix $\mathbf{V}$ obtained from the covariance matrix $\mathbf{C}_{\mathbf{X}^N}$ of the noise part. The transformed data is given by

$$\mathbf{V}^T\mathbf{X} = \mathbf{V}^T\mathbf{X}^S + \mathbf{V}^T\mathbf{X}^N, \tag{5.4}$$

where the transformed covariance matrix is $\mathbf{V}^T\mathbf{C}_\mathbf{X}\mathbf{V}$. The transformed noise part $\mathbf{V}^T\mathbf{X}^N$ is white, thus $\mathbf{V}^T\mathbf{C}_{\mathbf{X}^N}\mathbf{V} = \mathbf{I}$. The second transformation consists of an ordinary PCA, given by

$$\mathbf{E}^T\mathbf{V}^T\mathbf{C}_\mathbf{X}\mathbf{V}\mathbf{E} = \mathbf{\Lambda}, \tag{5.5}$$

where $\mathbf{E}$ is the matrix of orthogonal eigenvectors and $\mathbf{\Lambda}$ the matrix of the corresponding eigenvalues, sorted by decreasing order.
The NAPCA estimated subspace basis is given by [Lee90, Wüb07]

$$\hat{\mathbf{B}} = (\mathbf{V}^T)^{-1}\tilde{\mathbf{E}}, \tag{5.6}$$

where $\tilde{\mathbf{E}}$ is the matrix that contains the first M_e eigenvectors.The corresponding signals are estimated by linearly projecting the data onto the subspace basis without inverting the first NAPCA transformation, which may be expressed as

$$\hat{\mathbf{S}} = \tilde{\mathbf{E}}^T\mathbf{V}^T\mathbf{X}. \tag{5.7}$$

Figure 5.2 illustrates the working mechanism of NAPCA. In (a) the signal part and the noise part of the signals are depicted. An ordinary PCA would have the principal components in the direction of the noise. Transforming the data with the first step of NAPCA whitens the noise (b) and transforms the signal part. The principal component of a PCA of the transformed data yields

the mixing vector of the evoked signal.

A virtual experiment was set up using a two-dimensional evoked signal subspace. 100 randomly placed ECDs were used to model unwanted signals with different energies. The mean SNR of the simulated MEG data was dependent on the realization of the experiment and ranged between -10 dB and -15 dB. Sensor noise with 30 dB simulated data to sensor noise ratio was added. 100 runs were evaluated, each having the same evoked signal forms and signal power, while the location of the evoked signals in the sphere was chosen at random. For each realization, 100 trials with 400 samples were considered and ICA followed by single-trial SNR based grouping [Lem06, Koh10b] as well as NAPCA were run.

The performance depicted in Fig. 5.3 is evaluated in terms of SNR of recovered single-trials [Lem06]. ICA followed by grouping is given a slight advantage as this method aims at maximizing the very quantity that is used to evaluate the performance, i.e. SNR. However, Fig. 5.3 suggests higher potential of using NAPCA. It suggests that PCA is not suited to recover the subspace. ICA followed by grouping yields better results, while NAPCA enjoys best performance.
Hence, in the forthcoming, NAPCA will be run as a preprocessing method before decomposition in order to extract the evoked subspace. Using this procedure, dimensionality is reduced and so is the complexity of the separation task. For the methods based on structured model, this step is essential in order to assure that these models can be applied.

5.2 Bilinear methods beyond ICA

In this section, methods that go beyond ICA shall be discussed. All of these have in common that the dependent source problem is tackled, while having a close link to the established ICA methods, i.e. while including the independence scenario. They shall be termed here as *reference* methods. Next to ICA, novel approaches shall be compared to these selected reference methods in order to assess their value.

Spatial ICA (sICA)

McKeown et al. invented spatial ICA (sICA) [McK98] in order to apply this technique for fMRI data (For more details refer also to [Cal09]). In contrast to ordinary ICA, the columns of the mixing matrix $\mathbf{A}$ are assumed to be mutually independent, while the sources $\mathbf{S}$ can be dependent. Spatial ICA is easily computed using ordinary ICA but feeding the data in a transposed manner. In particular, with

$$\mathbf{X}^T = (\mathbf{AS})^T = \mathbf{S}^T\mathbf{A}^T, \tag{5.8}$$

the rows of the signal matrix become the mixing matrix, while the columns of the mixing matrix become the signal rows to be separated. Hence, in spatial ICA the mixing matrix, i.e. field patterns, are assumed to be independent.
For evoked MEG data, the number of samples in signal matrix $\mathbf{S}$ is usually much larger than the

number of samples in the mixing matrix $\mathbf{A}$. However, using measuring systems with more than 100 channels, spatial ICA may become a potential candidate for separating dependent source signals.
Spatial ICA has been combined with temporal ICA in [Sto02, The05] to give more robust estimates. For this, both the signal rows as well as the mixing columns are assumed to be independent. Running spatial ICA and temporal ICA in sequence gives estimates that can be compared to check the plausibility of the results. If both estimates show identical results a successful decomposition is most likely. Here, spatial ICA shall serve as the first reference method.

Subband decomposition ICA - Tanaka's approach (SDICA 1)

The next four methods aim at using ICA after preprocessing, such that the underlying sources are made less dependent. Subband decomposition ICA (SDICA1) [Tan04] uses a bank of filters and assumes that at least two source subbands are less dependent than the unfiltered dependent sources.

In particular, SDICA1 considers each source signal $\mathbf{s}$ as the sum of L subsignals, given by

$$\mathbf{s} = \sum_i \mathbf{s}_i, \quad \mathrm{T}_i(\mathbf{s}) = \mathbf{s}_i. \tag{5.9}$$

where $i = 1, ..., L$ and $\mathrm{T}_i(\cdot)$ is a linear transformation to obtain the subsignal corresponding to the ith subband.

In the same way, the recorded signal at some channel $\mathbf{x}$ is decomposed into the sum of L subsignals such that

$$\mathbf{x} = \sum_i \mathbf{x}_i, \quad \mathrm{T}_i(\mathbf{x}) = \mathbf{x}_i. \tag{5.10}$$

As T is a linear operator, the ith recorded subsignals $\mathbf{X}_i$ are a linear mixture of the ith source subsignals $\mathbf{S}_i$, given by

$$\mathrm{T}_i(\mathbf{X}) = \mathrm{T}_i(\mathbf{AS}) = \mathbf{A}T_i(\mathbf{S}), \tag{5.11}$$

where T_i operate row-wise. Hence, a correct ICA unmixing matrix recovered from the ith data subband applies also to the original data as

$$\sum_i \mathrm{T}_i(\mathbf{X}) = \mathbf{A} \sum_i \mathrm{T}_i(\mathbf{S}). \tag{5.12}$$

A filter bank with 20 subbands from 0 Hz to 80 Hz is used. Chebycheff 2 filters are used having a stopband of 90 dB. Subsequently, ICA is run on all L subbands of the recorded data. In order to select a data subband with least dependent sources, Tanaka et al. [Tan04] proposed the use of a modified unmixing error as follows. The unmixing error given by Eq. 3.45 is in need of the 'original' mixing matrix $\mathbf{A}$. This matrix is replaced by the inverse of a subband ICA unmixing

matrix $\mathbf{W}_0$ yielding $\tilde{\mathbf{A}} = (\mathbf{W}_0)^{-1}$. The unmixing matrix $\mathbf{W}$ is obtained from another subband ICA unmixing matrix. By plugging

$$\tilde{\mathbf{G}} = \tilde{\mathbf{A}}\mathbf{W} \tag{5.13}$$

into Eq. 3.45, *blind* Amari unmixing error is obtained. If two subbands have independent source subsignals blind amari error will be zero as the two unmixing matrices should be almost equal. Hence, the two data subbands that correspond to least blind unmixing error are selected and yield the unmixing matrix $\mathbf{W}$ that is applied to the original data.

SDICA1 is appealing because of its automated choice of the least dependent subbands. However, very small frequency bands are similar to sinusoidal signals differing only in phase. Such signals are not independent. Hence, a too high number of subbands will yield dependent source subsignals. In contrast, a too low number of subbands may not alter the original dependent sources enough. Hence, the number of subbands and the cut-off frequency are parameters that have to be chosen wisely. Nevertheless, this method shall be considered as a general reference method as no fine-tuning of these parameters is done.

Subband decomposition ICA - Kopriva's approach (SDICA 2)

Kopriva [Kop08] proposed to use wavelet tree decomposition for subband decomposition ICA (SDICA 2). More importantly, band selection is based on the assumption, that the data subband with smallest mutual information corresponds to the source subband with smallest mutual information. If this assumption holds, it is possible to find the most independent source subband by estimating the mutual information of all data subband and choosing the least dependent among them. Subsequent use of ICA of that subband assures lowest mutual information among the source subbands and, hence, best separation performance.
However, the proof in [Kop08] cannot be correct in general [Lin09], which shall be exemplified in the following.

Let the data $\mathbf{X}$ be given by two data subbands $\mathbf{X}_i$ and $\mathbf{X}_j$. The sources in the ith subband $\mathbf{S}_i$ shall be dependent, while the sources in the jth subband $\mathbf{S}_j$ shall be independent. With

$$\mathbf{X}_i = \mathbf{A}\mathbf{S}_i, \tag{5.14}$$
$$\mathbf{X}_j = \mathbf{A}\mathbf{S}_j, \tag{5.15}$$

the proof in [Kop08] claims that $\mathbf{X}_j$ is less dependent than $\mathbf{X}_i$, which gives the least dependent source subband, i.e. in our case the jth subband. However, this is not correct for at least one signal constellation. E.g., let the dependent sources in the ith subband be given by

$$\mathbf{S}_i = \mathbf{B}\tilde{\mathbf{S}}_i, \tag{5.16}$$

where $\tilde{\mathbf{S}}_i$ shall contain independent sources and $\mathbf{S}_i$ shall contain dependent sources due to the linear transformation $\mathbf{B}$. Furthermore, let the mixing matrix be given by

$$\mathbf{A} = \mathbf{B}^{-1}. \tag{5.17}$$

Mixing the sources using Eq. 5.14 together with Eq. 5.16 and 5.17 yields

$$\mathbf{X}_i = \mathbf{A}\mathbf{S}_i = \mathbf{A}\mathbf{B}\tilde{\mathbf{S}}_i = \tilde{\mathbf{S}}_i, \tag{5.18}$$
$$\mathbf{X}_j = \mathbf{A}\mathbf{S}_j = \mathbf{B}^{-1}\mathbf{S}_j. \tag{5.19}$$

Hence, the ith data subband $\mathbf{X}_i$ is independent, while the sources $\mathbf{S}_i$ are dependent. The jth data subband $\mathbf{X}_j$ is dependent, while the sources $\mathbf{S}_j$ are independent. It follows that choosing the ith data subband does not correspond to the most independent source subband.

Put into words, if the source signals are dependent because of a linear mixing then linear unmixing can make them independent. The corresponding data subband may become independent, while the source subband is dependent. Another source subband may be less dependent and the choice of the least dependent data subband may not be optimal.
However, one may argue that in most cases linear mixing makes the source subbands more dependent. The heuristic of choosing a data subband that is least dependent is still of interest. Indeed, SDICA 2 was shown to be successful using real world data in [Kop08] and shall serve as reference method here.

Highpass ICA (hpICA)

A more straight forward prefiltering concept is that of simply applying a highpass filter before using ICA [Cic03]. Indeed, a highpass often reduces dependencies. This may be understood considering that small time shifts in the lower frequency range only slightly change dependencies as few new sample combinations occur. In contrast, small time shifts in the higher frequency range introduce more new sample combinations yielding more independent evoked signals. As small time shifts often occur from trial to trial, a highpass filter shall give less dependent source signals.
From this point of view, highpass filtering and subsequent use of ICA (hpICA) generalizes ICA. Ad hoc, a 12th order highpass butterworth filter with 80 Hz cut-off frequency is utilized without any fine-tuning.

Innovation ICA (innoICA)

A closely related approach to hpICA is ICA of innovations (innoICA). Particularly, Hyvärinen proposed in [Hyv98] to model the recordings as a first order AR process. The difference between the AR modeled data and the original data is termed innovations. As the AR operation is linear, the innovations of the sources are mixed to give the innovations of the recordings. It is argued that source innovations are more independent than the original sources, as they are not explainable

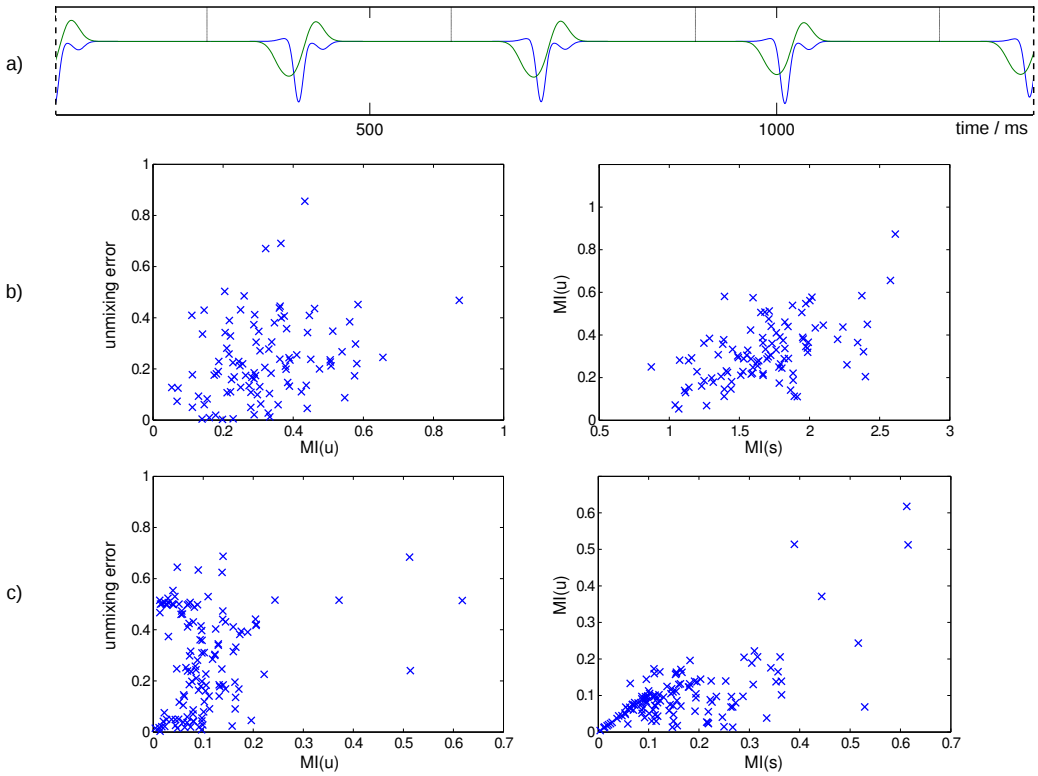


Figure 5.4: Binned mutual information [Kra04] of dependent sources and estimates using random mixing. a) Typical dependent source signals. b) Results from ICA based on binned mutual information for 50 dependent signals. c) Result from FastICA of 50 filtered versions of a single mixture using 50 different filters.
Mutual information of the original signals is often changed by ICA. Low mutual information of ICA estimates can correspond to high unmixing error. Hence, ICA can fail even when the ICA estimates are almost independent. In this case, the dependent subspace cannot be found by using the mutual information of this ICA estimate. Thus, using mutual information of ICA estimates for subspace analysis or subband ICA is questionable.

by the AR model. A 5th order AR model shall be used to feed the innovations to ICA as the last reference method considered.

5.3 On mutual information of ICA results

In the last two sections, techniques were discussed that use mutual information of ICA results to group dependent components for subspace analysis as well as for finding least dependent subbands to separate dependent sources. The latter approach may be exemplified further by the ideas in [Zha06a, Zha06b], where the authors employ an adaptive filter that is adjusted such that ICA

yields most independent estimates. Despite these straight forward solutions, the use of mutual information of ICA estimates for subspace and independent component analysis is not without drawbacks and shall be discussed in the following.
Specifically, it was argued in section 4.1.1 that ICA can separate dependent data, if the original dependent sources are not changed by an ICA. However, ICA fails, if the original dependent sources are changed by ICA, i.e. some linear transformations lower some ICA contrast and the sources are mixed up. This has consequences on techniques that are based on evaluating mutual information of ICA results.
Let a crude ICA method be based on binned mutual information [Kra04]. Furthermore, let 50 dependent signals, such as depicted in Fig. 5.4 a), be mixed by 50 randomly chosen mixing matrices. Fig. 5.4 b) depicts binned mutual information of the 50 source signals and the 50 ICA estimated signals, respectively. If the data were correctly unmixed, sources and estimates would have the same binned mutual information. However, it can be clearly observed from Fig. 5.4 b) right, that the binned mutual information of the original signals is lowered by ICA in all cases. This is not surprising, as lowest binned mutual information is seeked. Fig. 5.4 b) left, depicts mutual information of estimates and the corresponding unmixing error. No trend indicates that higher mutual information of the estimates is linked to lower performance of ICA or vice versa. Contrarily, very low mutual information of ICA estimates can have high separation errors associated.
Along this line, let a second experiment be based on FastICA. A single source signal pair as depicted in Fig. 5.4 a) shall be used and a selection of 50 different filters shall give 50 different subsignals. The logic of some prefiltering techniques is to design a filter such that ICA performs best, evaluated by the level of mutual information of the estimates. However, 5.4 c) cannot suggest a general success of this technique using the depicted signal pair. Specifically, Fig. 5.4 c) right shows that the mutual information of the original signals is often changed by the use of ICA. Fig. 5.4 c) left shows that independent results vary largely between low and high unmixing error.
As a consequence, using mutual information of ICA estimates for unmixing or for grouping is questionable. Low mutual information of the estimates can be linked to high unmixing errors. Hence, ICA fails but a grouping algorithm is likely to miss these sources for the dependent subspace.
Of course, often, mutual information of ICA outputs with underlying dependent signals will not be zero and ICA outputs from exactly independent signal will have most likely lower mutual information. However, to find a powerful test statistic may be difficult as real world signals often are not completely independent or the sample size and evaluation technique leads to deviation from zero mutual information. For subspace analysis, NAPCA may be more useful. For separation, data models beyond the bilinear shall prove useful and are considered in the forthcoming sections.

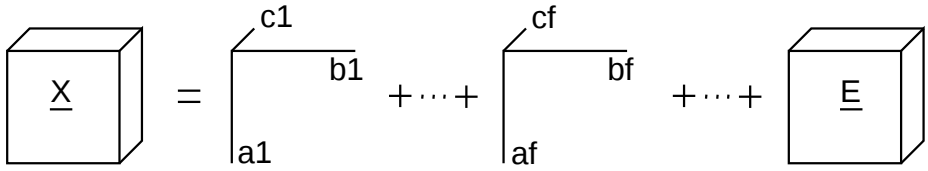


Figure 5.5: Graphical representation of CP decomposition using three-way arranged MEG data. F rank-one tensors represent F sources that are assumed to underly the MEG data. The residual is a three-way tensor of same size as the MEG data tensor.

5.4 Beyond the bilinear model

In order to tackle the problem of separating dependent components, a selection of reference methods were presented that try to generalize ICA in that sense. Many more methods exist that assume various objective functions next to independence. All these methods have in common that they need some assumption in order to render the bilinear mixing model essentially unique. This section shall discuss alternative models next to the bilinear model. In order to have an essentially unique model, the degrees of freedom of the bilinear model need to be reduced. A model that is not in need of the independence assumption must come at the prize of being more problem specific, i.e. to model more features of evoked MEG data than the bilinear model does. However, it must be general enough to cover all typical evoked signal settings.

5.4.1 Evoked data and trilinearity

Evoked MEG data are structured not only in time and space but also due to its repeating trials. By rearranging the evoked data corresponding to space, time and trial, a threeway data tensor can represent the evoked data recordings. In order to explain this data tensor by a mixture of evoked sources tensors, a simple evoked signal model that differs from Eq. 2.2 shall be assumed for now, which ignores latency shifts [Möc88, Fie91]. Specifically, the sources are assumed to remain at fixed physical locations, having fixed temporal waveforms differing only in amplitudes from trial to trial. Hence, for the fth source, the spatial profile $\mathbf{A}_f$ is stationary and the signal profile $\mathbf{S}_f$ is the same in every trial having just different amplitudes. This may be expressed by scalar multiplication of the signal profile $\mathbf{S}_f$ by the eth entry of an amplitude profile D_{ef} associated to the fth source. Furthermore, in each trial, the signal $D_{ef}\mathbf{S}_f$ is projected by scalar multiplication to every sensor channel, given by the spatial profile $\mathbf{A}_f$. The sum of $f = 1...F$ sources finally gives the mixture, which may be expressed as

$$X_{cte} = \sum_f A_{cf} S_{tf} D_{ef} + E_{cte}, \tag{5.20}$$

where the space modality is indexed by c, the time modality is indexed by t and the trial modality is indexed by e.
As only addition and scalar multiplication is involved, the generative model in Eq. 5.20 is *trilinear*, i.e. linear in $\mathbf{A}$, $\mathbf{S}$ and $\mathbf{D}$. Note, however that the trilinear model for evoked MEG data may

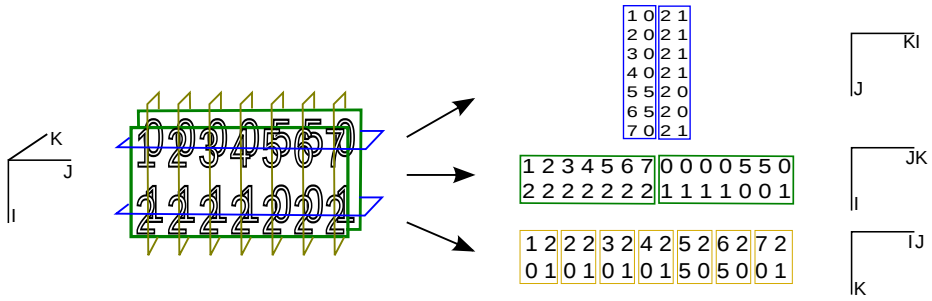


Figure 5.6: Unfolding concept demonstrated for a threeway array of dimensions $I \times J \times K$. Three different twoway matrices are expressed by unfolding in order to process the CP algorithm with ordinary twoway matrix manipulations.

be too simple as latency shifts are not modeled. The trilinear model can be identified essentially uniquely, which is a strong property, discussed in the following.

5.4.2 Identification

As discussed in Sec. 3.2, the bilinear model $\mathbf{X} = \mathbf{AS}$ is inherently non unique, as

$$\mathbf{X} = \mathbf{AS} = \mathbf{AUVS}. \tag{5.21}$$

holds true for any regular matrices with $\mathbf{U} = \mathbf{V}^{-1}$. Hence, infinitely many solutions to Eq. 5.21 exist; all represent the data equally well and an additional assumption, such as independence, is needed to render the bilinear model essentially unique.

Let the same data be represented by a trilinear model. Specifically, let the eth slab of the threeway arranged data tensor $\underline{\mathbf{X}}$ in Eq. 5.20, i.e. the eth twoway matrix cut from the threeway data tensor $\underline{\mathbf{X}}$, be expressed as

$$\mathbf{X}_e = \mathbf{A}\ \mathrm{diag}(\mathbf{D}_e)\mathbf{S}^T. \tag{5.22}$$

As only one slab is considered, standard twoway matrix notation can be used. Let all matrices have full rank. It shall be proven that $\mathbf{A}$ and $\mathbf{S}$ are unique, given $(\mathbf{D}_e)$. For this, the regular matrices $\mathbf{U}$ and $\mathbf{V}$ again are introduced into Eq. 5.22 yielding

$$\mathbf{X}_e = \mathbf{A}\ \mathrm{diag}(\mathbf{D}_e)\mathbf{S}^T = \mathbf{AU}\ \mathrm{diag}(\mathbf{D}_e)\mathbf{VS}^T. \tag{5.23}$$

Eq. 5.22 holds true for any regular matrices $\mathbf{U}$ and $\mathbf{V}$ that meet the identity

$$\mathrm{diag}(\mathbf{D}_e) = \mathbf{U}\ \mathrm{diag}(\mathbf{D}_e)\mathbf{V}. \tag{5.24}$$

With $\mathbf{U} = \mathrm{diag}(\mathbf{D}_e)\mathbf{V}^{-1}\mathrm{diag}(\mathbf{D}_e)^{-1}$ infinitely many solutions to Eq. 5.21 exist only, when e is fixed to one specific trial, say e_0. If more than one trial e is considered, the identity Eq. 5.24 needs to be reformulated as

$$\begin{aligned} \text{diag}(\mathbf{D}_e) &= \mathbf{U}\,\text{diag}(\mathbf{D}_e)\mathbf{V}, && (5.25)\\ &= \text{diag}(\mathbf{D}_e)\mathbf{U}\mathbf{V} && (5.26) \end{aligned}$$

which holds true only for diagonal matrices $\mathbf{U}$ and $\mathbf{V}$ and if $\mathbf{U} = \mathbf{V}^{-1}$. It follows that, given $(\mathbf{D}_e)$, Eq. 5.22 is unique.

To prove that the complete trilinear model is unique, let Eq. 5.23 be reformulated as [Mor08]

$$\begin{aligned} \mathbf{X}_e &= \mathbf{A}\,\text{diag}(\mathbf{D}_e)\mathbf{S}^T, && (5.27)\\ &= \mathbf{A}\mathbf{U}[\mathbf{U}^{-1}\,\text{diag}(\mathbf{D}_e)\mathbf{V}]\mathbf{V}^{-1}\mathbf{S}^T, && (5.28)\\ &= \tilde{\mathbf{A}}\,\text{diag}(\tilde{\mathbf{D}}_e)\tilde{\mathbf{S}}^T. && (5.29) \end{aligned}$$

As the term in square brackets needs to stay diagonal (constrained by the trilinear model), the matrices $\mathbf{U}$ and $\mathbf{V}$ can only be of permuted diagonal form. In particular, Eq. 5.29 imposes

$$\mathbf{U} = \mathbf{P}\mathbf{D} \quad \text{with } \mathbf{U}^{-1} = \mathbf{D}_0^{-1}\mathbf{P}^{-1} \tag{5.30}$$

and

$$\mathbf{V} = \mathbf{P}\mathbf{D}_1 \quad \text{with } \mathbf{V}^{-1} = \mathbf{D}_1^{-1}\mathbf{P}^{-1} \tag{5.31}$$

in order to be valid. Note that the permutation matrix is restricted to be the same for both matrices matrices $\mathbf{U}$ and $\mathbf{V}$ to ensure a diagonal matrix in the square brackets. However, $\mathbf{D}_0$ and $\mathbf{D}_1$ are allowed to differ. This states that energy can be held by the amplitudes, signals or mixing, while permutation of sources has to be done for all three accordingly. This proves that the trilinear model is unique up to the indeterminacies sign, norm and order [Har72, Möc88].
The assumption of full matrices rank can be further relaxed. Kruskal showed that the trilinear model is essentially unique if

$$k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2F + 2, \tag{5.32}$$

where F denotes the rank of the data tensor and k is the Kruskal rank, i.e. the smallest subset of linearly independent columns of the associated matrix with $k_{\mathbf{A}} \leq \text{rank}(\mathbf{A})$ [Kru77].

5.4.3 The CP model

Trilinear tensor decomposition was originally introduced by Hitchcock in 1927 [Hit27], but became popular later under the name canonical decomposition (CANDECOMP) [Car70] and parallel factor analysis (PARAFAC) [Har70]. More recently, it has been named as CP in order to account for all three works [Kie00]. Möcks [Möc88] reinvented CP in the context of brain imaging under the name topographic components model. The CP model was first proposed by Field et al. in the context of evoked brain sources [Fie91]. However, problems arose and CP on wavelet

transformed brain data was introduced in [Miw04, Mor05].
CP stands for fitting the trilinear model of Eq. 5.20 to structured threeway data in the least squares sense. In particular, the threeway data tensor is decomposed into a sum of rank-one tensors. A rank-one tensor is the outer product of vectors, i.e. each element of the rank-one tensor equals the multiplication of scalars taken from these vectors with corresponding indices. The CP decomposition structure is depicted in Fig. 5.5.
The least squares objective function is given by

$$\mathrm{E} = ||\underline{\mathbf{X}} - \underline{\hat{\mathbf{X}}}||_F^2, \tag{5.33}$$

where $\hat{\mathbf{X}}$ denotes the trilinear CP model, i.e. Eq. 5.20 without noise yielding

$$\hat{X}_{cte} = \sum_f A_{cf} S_{tf} D_{ef}. \tag{5.34}$$

Optimization of the least squares objective using Eq. 5.33 involves estimation of the matrices $\mathbf{A}$, $\mathbf{S}$, $\mathbf{D}$ that are dependent on each other and amounts to a difficult nonlinear problem [Bro98]. As a consequence, the CP model is commonly estimated by using alternating least squares (ALS) optimization. ALS optimizes one subset of parameters given the remaining subsets. Linear least squares (LLS) provide the solution to one subset given the remaining subsets and the estimation is iterated until the model fit is not further improved. A fortunate feature of ALS is that in each step, LLS forces the objective to decrease or to remain unchanged, which yields convergence [Bro98]. In particular the estimation of LLS amounts to $\mathbf{x} = \mathbf{A}\mathbf{s}$, where a data vector $\mathbf{x}$ lies in the recording space of dimensionality equal to the number of channels. However, the columnspace of the mixing matrix $\mathbf{A}$ most often is mostly a lower dimensional subspace of the recording space. i.e. less sources than channels $\mathbf{s}$ give rise to less columns in $\mathbf{A}$ than channels. Hence, only if the data vector $\mathbf{x}$ lies in the columnspace of $\mathbf{A}$, a solution exists and the source vector $\mathbf{s}$ is readily obtained by $\mathbf{s} = \mathbf{A}^{-1}\mathbf{x}$. Due to noise, the data vector $\mathbf{x}$ will most likely not be comprised by the columnspace of $\mathbf{A}$.
Hence, linear least squares amount to find a solution to $\mathbf{s}$, such that the error vector between the modeled data and the true data has smallest squared norm, i.e. least sum of squared error. This is the case, when the error vector $\mathbf{x} - \mathbf{A}\mathbf{s}$ is orthogonal to the columnspace, i.e. to all modeled data vectors $\mathbf{A}\mathbf{s}$, as the squared norm is assured to be minimal by geometry [Str80]. It follows that

$$\begin{aligned}
(\mathbf{A}\mathbf{s})^T(\mathbf{x} - \mathbf{A}\mathbf{s}) &= 0 && (5.35)\\
\mathbf{s}^T\mathbf{A}^T\mathbf{x} - \mathbf{s}^T\mathbf{A}^T\mathbf{A}\mathbf{s} &= 0 && (5.36)\\
\mathbf{A}^T\mathbf{A}\mathbf{s} &= \mathbf{A}^T\mathbf{x} && (5.37)\\
\mathbf{s} &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{x} && (5.38)\\
\mathbf{s} &= (\mathbf{A})^{\#}\mathbf{x} && (5.39)
\end{aligned}$$

where Eq. 5.39 yields the optimal least squares solution and $(\mathbf{A})^{\#}$ denotes the pseudo-inverse of the mixing matrix $\mathbf{A}$ equal to $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$.
In order use LLS for ALS, the tensor must be rearranged such that twoway matrix operations can

be performed. Tensor unfolding is used to arrange the data tensor and the Khatri-Rao product is used to model rank-one tensors in twoway matrix notation [Kol09]. In particular, a threeway tensor can be rearranged, by taking appropriate slices of the tensor in sequence. The process is referred to as unfolding and is depicted in Fig. 5.6. A threeway data representation has six different twoway matrices equivalents, however for performing ALS via LLS, only three are needed.
The rank-one source tensors, which are depicted in Fig. 5.5, are also threeway tensors. These can be given in twoway matrix notation using the Khatri-Rao product, which is given by

$$[\mathbf{D} \odot \mathbf{S}] = [\mathbf{D}_1 \otimes \mathbf{S}_1 \; \mathbf{D}_2 \otimes \mathbf{S}_2 \; \cdots \; \mathbf{D}_F \otimes \mathbf{S}_F], \tag{5.40}$$

i.e. which equals column-wise Kronecker product of the matrices involved [Rao71, Bro98].

For convenience, let an example clarify these notations. Let the matrix of amplitudes be given by

$$\mathbf{D} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 8 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \tag{5.41}$$

so that their Khatri-Rao product amounts to

$$\mathbf{D} \odot \mathbf{S} = \left[\; \mathbf{D}_1 \otimes \mathbf{S}_1 \; \mathbf{D}_2 \otimes \mathbf{S}_2 \; \mathbf{D}_3 \otimes \mathbf{S}_3 \; \right] = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 2 \\ 0 & -4 & 0 \\ 2 & 4 & 0 \\ 2 & -4 & 0 \\ 0 & 0 & 8 \end{bmatrix} \tag{5.42}$$

Let further the mixing matrix be given by

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \tag{5.43}$$

In the limit of no noise, the unfolded data tensor $X_{(1)}$, equals the mixture of the underlying rank-one source tensors, expressed using the Khatri-Rao product, yielding

$$\mathbf{X}_{(1)} = \mathbf{A}(\mathbf{D} \odot \mathbf{S})^T \tag{5.44}$$

$$\left[\begin{array}{cccc|cccc} 0 & 1 & 1 & 0 & 0 & 2 & 2 & 0 \\ -1 & 1 & -1 & 2 & -4 & 4 & -4 & 8 \\ -1 & 1 & -1 & 4 & -4 & 4 & -4 & 16 \end{array}\right] = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{array}\right] \left[\begin{array}{cc} 0 & -1 & 0 \\ 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 2 \\ 0 & -4 & 0 \\ 2 & 4 & 0 \\ 2 & -4 & 0 \\ 0 & 0 & 8 \end{array}\right]^T \tag{5.45}$$

With these notations at hand, the CP model given in Eq. 5.20 may be rewritten in three equal versions, given by

$$\mathbf{X}_{(1)} = \mathbf{A}(\mathbf{D} \odot \mathbf{S})^T + \mathbf{E}_{(1)}, \tag{5.46}$$
$$\mathbf{X}_{(2)} = \mathbf{S}(\mathbf{D} \odot \mathbf{A})^T + \mathbf{E}_{(2)}, \tag{5.47}$$
$$\mathbf{X}_{(3)} = \mathbf{D}(\mathbf{A} \odot \mathbf{S})^T + \mathbf{E}_{(3)}. \tag{5.48}$$

The threeway evoked MEG data matrix $\underline{\mathbf{X}} \in \mathbb{R}^{C \times T \times E}$ is unfolded, by taking appropriate slices of the data tensor $\underline{\mathbf{X}}$ in sequence, yielding regular twoway matrices, where $\mathbf{X}_{(1)} \in \mathbb{R}^{C \times TE}$, $\mathbf{X}_{(2)} \in \mathbb{R}^{T \times CE}$ and $\mathbf{X}_{(3)} \in \mathbb{R}^{E \times TC}$ and C, T and E being the number of channels, timepoints and trials, respectively. The noise tensor $\underline{\mathbf{E}}$ is unfolded likewise. The Khatri-Rao product $\mathbf{D} \odot \mathbf{S}$ gives a twoway matrix. Left multiplying the spatial matrix $\mathbf{A}$ and adding the unfolded noise tensor $\mathbf{E}_{(1)}$ equals the unfolded data $\mathbf{X}_{(1)}$, i.e. is no more than a different representation of the three way data in long two-way matrix notation.
The ALS procedure, may now be expressed as

$$\mathbf{A} \leftarrow \mathbf{X}_{(1)}(\mathbf{D} \odot \mathbf{S})^{T\#}, , \tag{5.49}$$
$$\mathbf{S} \leftarrow \mathbf{X}_{(2)}(\mathbf{D} \odot \mathbf{A})^{T\#}, , \tag{5.50}$$
$$\mathbf{D} \leftarrow \mathbf{X}_{(3)}(\mathbf{A} \odot \mathbf{S})^{T\#}, \tag{5.51}$$

where $T\#$ denotes the pseudoinverse, i.e. ordinary LLS is involved. The ALS procedure to fit the CP model to the recordings is the alternation through Eq. 5.49 to Eq. 5.51. In each step the residual cannot increase. Due to the uniqueness results stated in the last section, only the original components yield minimal residual and the components are identified by this procedure if the model is correct. For this, the number of underlying components have to be known. They can be estimated by techniques described in [Bro03] but are assumed known throughout this work.

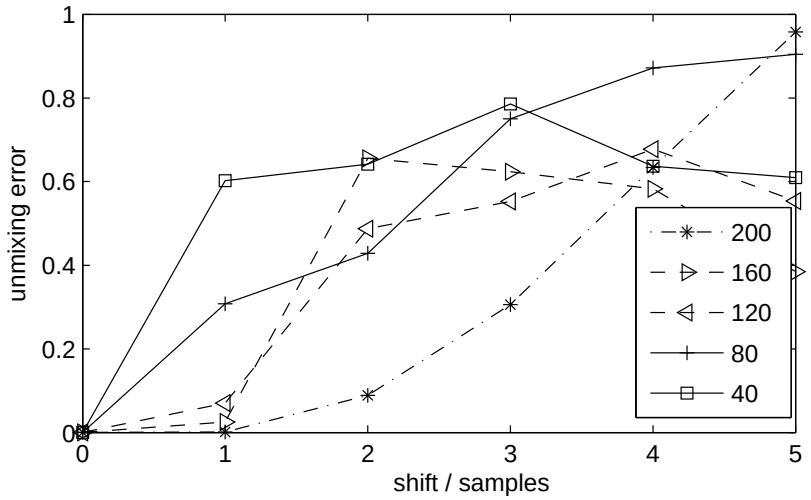


Figure 5.7: Robustness analysis of CP against shifts given different signals shapes. A bell shaped curve and its derivative were used. Both have a temporal support of $200, 160, 120, 80$ and 40 samples. Already for a support of 200 samples, shifts of up to 3 samples lead to problems for the CP model. For smaller temporal supports, i.e. higher frequency contents, CP has troubles with shifts of only 1 sample. This shows that CP is not robust to shifts, dependent on the spectral content of the signal shapes.

5.4.4 Shifts can destroy trilinearity

Varying latencies over trials result in shifts in the time modality from trial to trial, violating the trilinear assumption. It is a reasonable question to ask how much shift is allowed in CP and when a shifted factor analysis is needed. This question cannot be answered in general without the knowledge of the signal shapes and its spectral content. For a signal with spectral content in the very low frequency range small shifts shall not violate the CP model assumptions too much. However the same latency shift scenario for signal shapes that have higher frequency contents may lead to a failure of the CP model identification.
This is exemplified in Fig. 5.7. Here, a bell shaped curve and its derivative were used with a temporal support of 200 samples duration down to 40 samples duration in steps of 40 samples. Hence, both the temporal support of signal was lowered and the frequency content was augmented. Fig. 5.7 suggests that already for a support of 200 samples, shifts of up to 3 samples lead to problems for the CP model. The same form in higher frequency band, i.e. with a smaller temporal support, leads to failure of the CP model already for shifts up to 1 sample. This is a remarkable result, as it shows that the CP is not robust to shifts when signals have limited temporal support with higher frequency contents. This situation can easily occur in evoked signals and, hence, the incorporation of shifts may improve results. Shifted factor analysis shall be discussed in the following.

5.5 Shifted factor analysis

Let evoked MEG data be modeled as a superposition of the magnetic fields induced by neuronal source signals. With each neuronal signal a particular spatial pattern of the induced magnetic field is associated. The single neuronal signals and hence their induced magnetic fields are subject to amplitude and latency variations within the epochs, while the form of a neuronal signal is assumed to be constant throughout all epochs. Let X_{cte} denote the MEG data obtained for the eth epoch in the cth channel at the time t after imposing the stimulus. According to our assumptions the MEG data can then be described by the SFA model [Har03, Hon03a, Hon03b]

$$X_{cte} = \sum_f A_{cf} S_f(t + T_{ef}) D_{ef} + N_{cte}, \tag{5.52}$$

where the matrix $\mathbf{A}$ encodes the spatial mapping from the neuronal source signals to the MEG channel, $\mathbf{T}$ the latency shifts and $\mathbf{D}$ the amplitude shifts. The fth neuronal signal form is denoted by $S_f(t)$. For notational convenience $S_{tf} = S_f(t)$ shall be used interchangeably. For instance, $A_{cf}S_{tf}$ is the magnetic signal at the cth channel evoked by the fth neuronal source at time t after stimulus presentation when no amplitude and latency variation is present. The term N_{cte} denotes noise which is assumed white and stationary.
The SFA model (5.52) is linear with respect to the spatial mapping $\mathbf{A}$ and amplitude variations $\mathbf{D}$, but it depends non-linearly on the neuronal signal forms $\mathbf{S}$ due to the assumed latency shifts $\mathbf{T}$. Assuming all matrices to be of full rank, Morup has proven that the SFA model (5.52) is essentially unique [Mor08], i.e. all of its unknowns (namely $\mathbf{A}$, $\mathbf{S}$, $\mathbf{T}$ and $\mathbf{D}$) can be determined essentially uniquely from the data X_{cte}. Hence, if the SFA model (5.52) applies for evoked MEG data, estimation of its unknowns reveals the evoked neuronal source signals.

However, the SFA model is not as easy to identify compared to CP. Due to the nonlinearity, estimation of its unknown is challenging. For instance, when applying least-squares estimation a nonlinear (and possibly non-convex) program has to be solved which may have several local solutions. There exist methods that are based on an exhaustive search of shifts [Har03, Hon03a, Hon03b] as well as a Bayesian method [Knu06] that aim at identifying the SFA model. However, these methods are numerically demanding, while only integer shifts are considered. In Sec. 5.4.4, trilinear modeling was shown to be sensitive to violations due to very small shifts, possibly non sample spaced. Hence, non-integer shifts shall be considered, i.e. shifts that may not be a multiple of the sample frequency. The first method that estimates the SFA model efficiently, while considering non-integer shifts, is frequency domain shifted factor analysis [Mor08]. It shall be introduced in the next section and considered as another reference method for the comparison study.

5.5.1 Frequency domain shifted factor analysis

Frequency domain shifted factor analysis (FDSFA) - originally named to as shifted CP (SCP)- [Mor08] is obtained by transforming the temporal mode of Eq. 5.52 into the frequency domain.

Shifts in the time domain correspond to linear phase shift in the frequency domain, using fast Fourier transformation, the transformation may be expressed as

$$S_f(t + T_{ef}) \circ\!\!-\!\!\bullet\, S_f(\nu)\text{exp}(-j2\pi\frac{\nu-1}{N}T_{ef}), \tag{5.53}$$

where ν, $-j$ and N denote frequency, imaginary unit and total number of samples of the temporal mode, respectively.
With Eq. 5.53, Eq. 5.52 is rewritten as

$$X_{c\nu e} = \sum_f A_{cf}S_{\nu f}D_{ef}\text{exp}(-j2\pi\frac{\nu-1}{N}T_{ef}) + N_{c\nu e}. \tag{5.54}$$

FDSFA aims at minimizing the objective

$$\sum_{c,\nu,e}\left(X_{c\nu e} - \sum_f A_{cf}S_{\nu f}D_{ef}\text{exp}(-j2\pi\frac{\nu-1}{N}T_{ef})\right)^2 \tag{5.55}$$

with respect to $\mathbf{S}, \mathbf{T}, \mathbf{A}$ and $\mathbf{D}$ using ALS. Specifically, let

$$S^{(e)}_{tf} = \text{IFFT}(S_{\nu,f}\text{exp}(-j2\pi\frac{\nu-1}{N}T_{ef})). \tag{5.56}$$

I.e. $\mathbf{S}^{(e)}$ denotes the matrix of source signals component wise shifted corresponding to the eth epoch. Let further

$$Z_{t+e(N-1),f} = D_{e,f}S^{(e)}_{tf}, \tag{5.57}$$

i.e. $\mathbf{Z}$ denotes the Khatri-Rao product between the matrix containing the amplitudes for each epoch and component and the source signal matrices containing the shifted source signals for each epoch and component, respectively. With Eq. 5.56 and Eq. 5.57, Eq. 5.54 is reformulated as

$$\mathbf{X}_{(1)} = \mathbf{A}\mathbf{Z}^T + \mathbf{E}_{(1)}, \tag{5.58}$$

$$\mathbf{X}_{(2)_\nu} = \mathbf{S}_\nu((\mathbf{D}\circ\text{exp}(-j2\pi\frac{\nu-1}{N}\mathbf{T}))\odot\mathbf{A})^T + \mathbf{E}_{(2)_\nu}, \tag{5.59}$$

$$\mathbf{X}_{(3)_e} = \mathbf{D}_e(\mathbf{A}\odot\mathbf{S}^{(e)})^T + \mathbf{E}_{(3)_e}, \tag{5.60}$$

where $\mathbf{D}\circ\mathbf{T}$ denotes the Hadamard product and $\mathbf{D}\odot\mathbf{S}$ the Khatri-Rao product of $\mathbf{D}$ and $\mathbf{T}$, respectively. Again, all three equations are replicas of Eq. 5.54, being three different twoway matrix representations.

The LLS updating equations for $\mathbf{A}$, $\mathbf{S}$ and $\mathbf{D}$ can then be expressed as

$$\mathbf{A} \leftarrow \mathbf{X}_{(1)}\mathbf{Z}^{T\#}, \tag{5.61}$$

$$\mathbf{S}_\nu \leftarrow \mathbf{X}_{(2)_\nu}(\mathbf{D}\circ\text{exp}(-j2\pi\frac{\nu-1}{N}T)\odot\mathbf{A})^{T\#}, \tag{5.62}$$

$$\mathbf{D}_e \leftarrow \mathbf{X}_{(3)_e}(\mathbf{A}\odot\mathbf{S}^{(e)})^{T\#}. \tag{5.63}$$

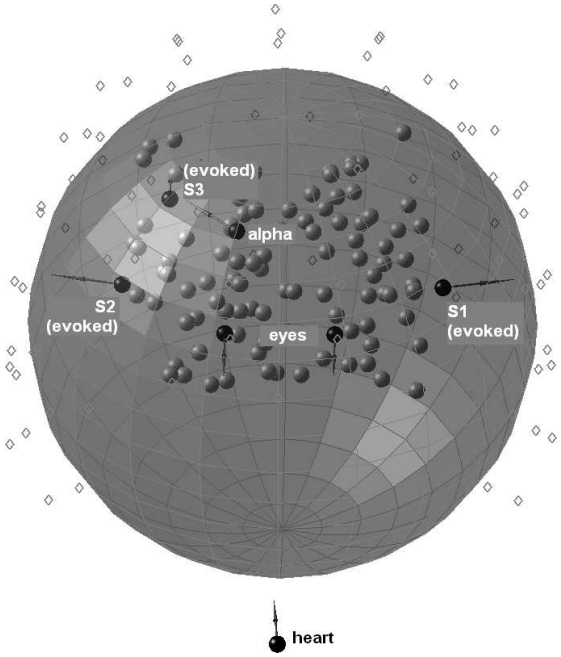


Figure 5.8: Sphere model of the human head with three evoked equivalent current dipoles (ECDs) (S1, S2, S3) and various background ECDs (alpha, eye, heart and noise). The gray spheres represent 100 background noise ECDs. Sensor locations (diamonds) correspond to a 93-channel MEG system.

By projecting out all but the f' source from the data at the eth epoch

$$R_{3^{(e)}_{f'}} = X_{(3)_e} - \sum_{f \neq f'} D_e f (A_f \odot S_f^{(e)})^T \tag{5.64}$$

the remaining signal is obtained. Maximum cross-correlation between $R_{3^{(e)}_{f'}}$ and $S_{f'}^{(e)}$ yield integer shift estimates. For the non-integer part, Morup et al. used a Newton-Raphson method starting from the integer estimate and optimized Eq. 5.55 with respect to $\mathbf{T}$ non-linearly. A detailed procedure can be found in the original works of Morup et al. [Mor08, Mor07].

5.6 Novel approach to shifted factor analysis in time domain

5.6.1 The 'infinite' CP model

Motivated by [Möc86, Fie91], a Taylor series expansion of the neuronal signals will be employed and it will be demonstrated that the non-trilinear SFA model can be transformed into a trilinear CP model that has infinitely many components.

By inserting a Taylor series expansion of the neuronal signals according to

$$S_f(t + T_{ef}) = \sum_m \frac{S_{tf}^{(m)}}{m!} T_{ef}^{\ m} \tag{5.65}$$

into Eq. 5.52 yields

$$X_{cte} = \sum_f \sum_m \frac{1}{m!} A_{cf} S_{tf}^{(m)} D_{ef} T_{ef}^{\ m} + N_{cte}, \tag{5.66}$$

which may be expressed as

$$X_{cte} = \sum_{f'} \tilde{A}_{cf'} \tilde{S}_{tf'} \tilde{D}_{ef'} + N_{cte}, \tag{5.67}$$

where f' encodes the tuple (m, f) in Eq. (5.66), $\tilde{A}_{cf'} = A_{cf}$, $\tilde{S}_{tf'} = S_{tf}^{(m)}$ and $\tilde{D}_{ef'} = \frac{1}{m!} D_{ef} T_{ef}^{\ m}$. The relation (5.67) shows that the SFA model (Eq. 2.2) is equivalent to a trilinear model with an infinite number of components.

5.6.2 Approximation of the model

Nevertheless, let the derivatives of the neuronal signals be approximated by linear functionals according to

$$S_{tf}^{(m)} = \sum_{t'} B_{tt'}^m S_{t'f}, \tag{5.68}$$

where the derivation matrix $\mathbf{B}$ may be set up, e.g., such that finite central differences are calculated. Inserting (5.68) into (5.66) gives

$$X_{cte} = \sum_f \sum_m \sum_{t'} \frac{1}{m!} A_{cf} B_{tt'}^m S_{t'f} D_{ef} T_{ef}^{\ m} + N_{cte}. \tag{5.69}$$

Since the matrices $\mathbf{B}^m$ are known, Eq. 5.69 contains as many unknowns as Eq. 5.54. For the application of the method, the Taylor series expansion has to be truncated. Note that the need for truncating the Taylor series expansion also offers a model flexibility. For instance, when latency shifts are small or even negligible, an order 0 Taylor series expansion would be sufficient and in this case the approximate SFA model (Eq. 5.69) is equivalent to a standard CP model. For larger latency shifts, on the other hand, a higher order Taylor series expansion is required and in this case model Eq. 5.69 is a good approximation to the SFA model (Eq. 5.54) For the results presented in Section 5, an ad-hoc order of 100 is applied. Such high order shall show the robustness as high order derivative are likely to have numerical problems. Furthermore, it is needed for extreme shifts. In reality, in turn, a much lower number should be sufficient. In the comparison study, a Taylor order of 10 is chosen, although good results have been obtained with an order 4 and for slow varying signals with an order down to 1.

5.6.3 Numerical implementation

Least squares estimation is proposed to determine the unknowns $\mathbf{A}, \mathbf{S}, \mathbf{T}, \mathbf{D}$ of the approximate SFA model (Eq. 5.69) obtained by truncating the Taylor series expansion. For this, the term

$$\sum_{c,t,e} \left(X_{cte} - \sum_{m} \sum_{f} \sum_{t'} \frac{1}{m!} A_{cf} B^m_{tt'} S_{t'f} D_{ef} T^m_{ef} \right)^2 \tag{5.70}$$

is minimized with respect to $\mathbf{A}, \mathbf{S}, \mathbf{T}, \mathbf{D}$. The unknowns $\mathbf{A}, \mathbf{S}, \mathbf{D}$ enter linearly in Eq. 5.69, and hence the linear least squares (LLS) techniques well-known in CP model identification [Bro97] are employed for their updates during the ALS procedure. The latencies $\mathbf{T}$ enter nonlinearly, and in each iteration of the ALS procedure they are updated by application of a Levenberg-Marquardt algorithm [Gil81].

CP techniques and notations are used to transfer the threeway model into standard twoway matrices. The data X_{cte} is unfolded, i.e. rearranged, yielding three different twoway representations $\mathbf{X}_{(i)}, i = 1...3$ [Bro97, p. 157]. For instance, the transposed c'th row of matrix $\mathbf{X}_{(1)}$ is obtained by concatenating the columns of the matrix $\mathbf{C}$ with elements $C_{te} = X_{cte}$. The same applies to the noise term N_{cte}. Using this notation, Eq. 5.67 can be expressed as

$$\mathbf{X}_{(1)} = \sum_{m} \mathbf{A}(\frac{\mathbf{D} \circ \mathbf{T}^m}{m!} \odot \mathbf{B}^m \mathbf{S})^T + \mathbf{N}_{(1)}, \tag{5.71}$$

$$\mathbf{X}_{(2)} = \sum_{m} \mathbf{B}^m \mathbf{S}(\frac{\mathbf{D} \circ \mathbf{T}^m}{m!} \odot \mathbf{A})^T + \mathbf{N}_{(2)}, \tag{5.72}$$

$$\mathbf{X}_{(3)} = \sum_{m} \frac{\mathbf{D} \circ \mathbf{T}^m}{m!} (\mathbf{A} \odot \mathbf{B}^m \mathbf{S})^T + \mathbf{N}_{(3)}, \tag{5.73}$$

where $\circ$ denotes Hadamard product and $\odot$ the Khatri-Rao product. All of these equations are just replicas of Eq. 5.69, i.e. three different twoway matrix representations. Note that the signal form matrix $\mathbf{S}$ is enclosed by two matrices in Eq. 5.72. The LLS updating equations for $\mathbf{A}$, $\mathbf{S}$ and $\mathbf{D}$ can then be expressed as

$$\mathbf{A} \leftarrow \mathbf{X}_{(1)}(\sum_{m}(\frac{\mathbf{D} \circ \mathbf{T}^m}{m!} \odot \mathbf{B}^m \mathbf{S})^T)^{\#}, \tag{5.74}$$

$$\mathbf{S} \leftarrow \text{mat}(\ (\sum_{m}(\frac{\mathbf{D} \circ \mathbf{T}^m}{m!} \odot \mathbf{A})^T \otimes \mathbf{B}^m)^{\#} \text{ vec } \mathbf{X}_{(2)}\ ), \tag{5.75}$$

$$\mathbf{D}_e \leftarrow \mathbf{X}_{(3)e}(\sum_{m} \text{diag}(\frac{\mathbf{T}^m_e}{m!})(\mathbf{A} \odot \mathbf{B}^m \mathbf{S})^T)^{\#}. \tag{5.76}$$

where $(\cdot)^{\#}$ denotes the Moore-Penrose inverse, and $\otimes$ is the Kronecker matrix product. Vec$(\mathbf{X})$ is a vector formed by concatenating the columns of the matrix $\mathbf{X}$ while mat$(\cdot)$ denotes the inverse operation. Diag$(\mathbf{v})$ is a diagonal matrix with diagonal elements equal to the vector $\mathbf{v}$. The subscript e denotes the e'th row of the corresponding matrix. For instance, $\mathbf{X}_{(3)e}$ is the eth row of the matrix $\mathbf{X}_{(3)}$. Eqs. 5.74 - 5.76 are the three LLS equations for our ALS procedure. Nonlinear minimization of Eq. 5.70 with respect to the latency shifts $\mathbf{T}$ then completes the proposed ALS

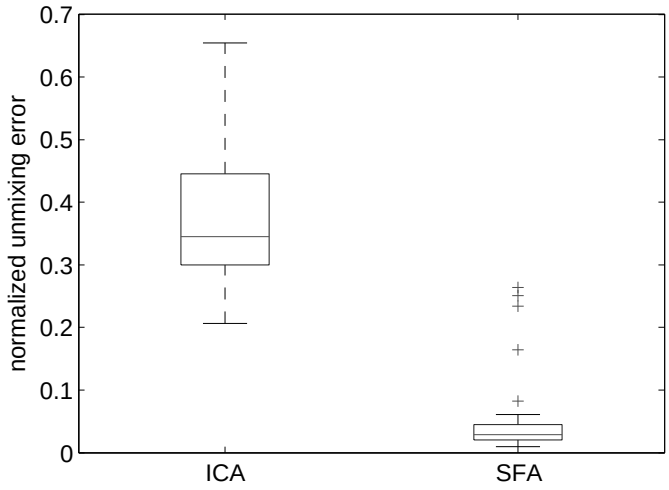


Figure 5.9: Unmixing performance for 50 different evoked signal settings that are dependent (maximum possible latency shifts $\delta = 6ms$).

procedure. Note that the nonlinear optimization step for updating the latencies allows for non-integer latencies to be determined.

When using a first-order Taylor series expansion in Eq. 5.65, the approximated SFA model Eq. 5.69 depends linearly also on the latencies $\mathbf{T}$. In this case, the nonlinear optimization step for updating the latencies can be replaced by a LLS step, given by

$$\mathbf{T}_e \leftarrow \tilde{\mathbf{X}}_{(3)e}(\text{diag}(\mathbf{D}_e)(\mathbf{A} \odot \mathbf{BS})^T)^{\#}, \tag{5.77}$$

where $\tilde{\mathbf{X}}_{(3)e} = \mathbf{X}_{(3)e} - \text{diag}(\mathbf{D}_e)(\mathbf{A} \odot \mathbf{S})^T$.
The ALS identification procedure was run using ICA for initialization of a single starting point. $\mathbf{A}$ was set to the ICA mixing estimate and $\mathbf{S}$ was set to the first epoch of the ICA signal estimate. $\mathbf{T}$ was initialized with integer shifts that maximized the cross-correlation of $\mathbf{S}$ with the corresponding epoch of the ICA signal estimate, while for $\mathbf{D}$ all values were set to 1. The iterations were terminated when the relative difference between the values of the cost function Eq. 5.70 in two subsequent iterations fell below the convergence level, which was set to 10^{-6}. The number of components was set to 3. The update order was $\mathbf{T}$, $\mathbf{D}$, $\mathbf{A}$, $\mathbf{S}$ while enhanced line search [Raj08] was used for acceleration. For further acceleration, arbitrary 5 channels and 5 epochs were utilized, and the ALS procedure was run with 200 time points until convergence or until 500 iterations were exceeded. Then, the ALS procedure was applied, considering either all epochs (and the selected 5 channels) or all channels (and the selected 5 epochs), while using the values for the unknowns obtained in the first step as the starting values. An appropriate choice of the Taylor order depends on the possible latency shift and the underlying signal form. In this work, the Taylor order was fixed at 100. Calculations were carried out on a compute server (8 Intel Xeon cores, 2.66 GHz, 128 GB RAM) with typical calculation times of 165 seconds, 712 seconds and 3243 seconds for the Taylor orders 1, 10 and 100, respectively.

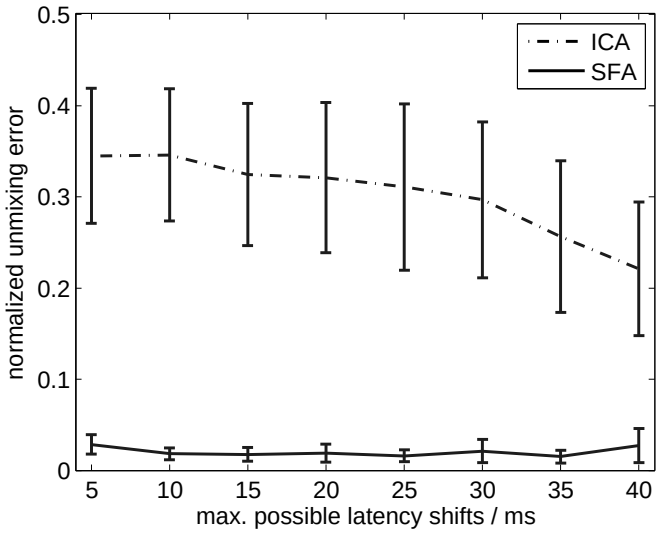


Figure 5.10: Unmixing performance of ICA versus the proposed SFA procedure for different dependency levels which are implicitly given by the maximal possible shifts.

5.6.4 Simulation results

SFA versus ICA

The simulated evoked MEG data were analyzed by the proposed procedure and the results were compared to those obtained by the Infomax ICA method [Bel95]. Assessment of blind source separation was done using the unmixing error. In addition, the latency estimates obtained by the proposed method were compared to those of an established single channel method [Jas99].
Fig. 5.9 shows the unmixing errors obtained for 50 different signal settings (amplitudes, latencies and signal forms) by the proposed procedure and the Infomax ICA method. The latencies were drawn from a uniform distribution on $[-\delta, \delta]$ with δ chosen equal to 5 ms. For this parameter δ, the resulting source signals show a high dependency. ICA fails as the signals are dependent, and the proposed procedure clearly outperforms ICA with respect to blind source separation here. But note that from the 50 different signal settings, 4 are apparently still not well separated. This can have several reasons. The convergence level or the choice and the order of approximating the derivatives may have been suboptimal. Another reason might be that the preprocessing step did not recover the signal subspace correctly. Furthermore, local minima could have been an issue. However, the results seem rather robust as no fine tuning was applied.

In a next step, the dependence of the source signals are controlled by varying the width of the uniform distribution $[-\delta, \delta]$ from which the latencies were drawn. In order to assess the consequences of different values of the maximal possible latency shifts δ, the mutual information between the source signals is determined as a function of δ. For δ = 5 ms, a high dependence can be observed, whereas for δ = 40 ms, the source signals appear to be only moderately dependent. Fig. 5.10 shows the unmixing errors in dependence on the parameter δ. In each case,

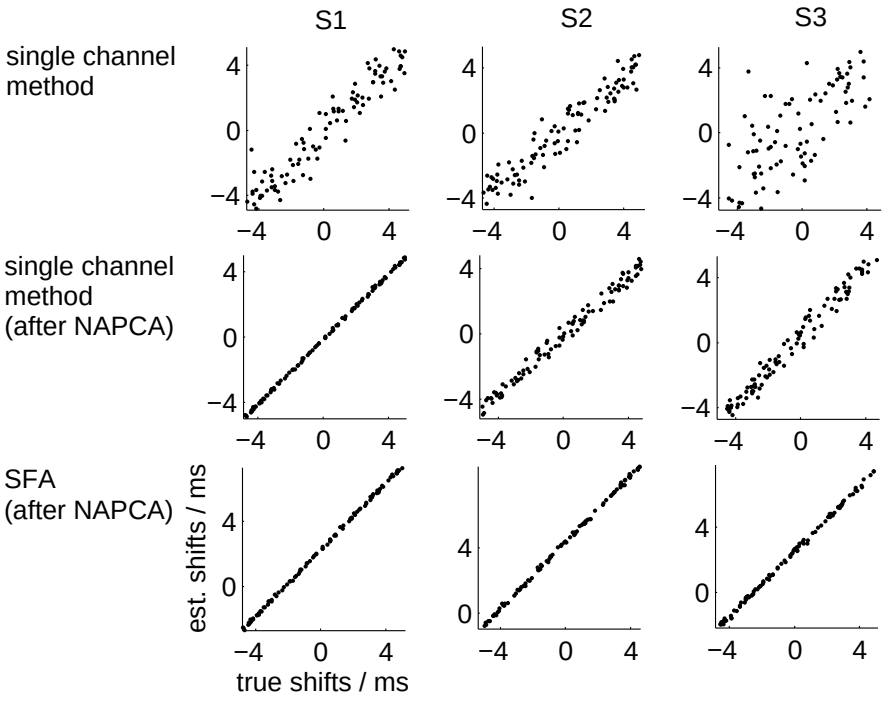


Figure 5.11: Latency estimation for maximal possible latency shifts δ=5 ms. For one setting of the three evoked signals, single channel maximum likelihood shift estimates [Jas99] are depicted in the first and second row using raw and NAPCA prefiltered data, respectively. The third row depicts typical latency shift estimates obtained by the proposed SFA procedure.

the results show the median together with the median absolute deviation from the unmixing errors obtained for 50 different signal settings. If the maximal possible latency shift was chosen to be greater than 100 ms ($\delta > 100$ ms) ICA was observed to work well. Fig. 5.10 shows more realistic shift settings choosing δ between 5 ms and 40 ms. ICA is observed to have difficulties even for moderate dependencies, when the normalized mutual information is less than 0.1, and fails for larger dependencies, while the proposed procedure appears to work well in all cases.

Unmixing error versus model fit

Throughout the whole work, the unmixing error is used as unmixing performance measure. In the context of least squares data fitting, however, the model fit is often looked at. Relative model fit may be expressed as

$$rMSE = \frac{||\underline{\mathbf{X}} - \underline{\hat{\mathbf{X}}}||_F^2}{||\underline{\mathbf{X}}||_F^2}, \tag{5.78}$$

where rMSE denotes sum of squared error normalized to sum of squared data. The fit value is dependent on the degrees of freedom of the utilized model. A model with as many degrees of

rMSE	2	4	6	8	10
PCA	4,82E-031	3,76E-029	4,01E-031	3,72E-030	1,53E-029
CP	3,06E-002	2,11E-002	1,74E-002	6,08E-003	4,61E-003
SFA	5,64E-004	6,61E-004	5,49E-004	1,01E-003	8,46E-004

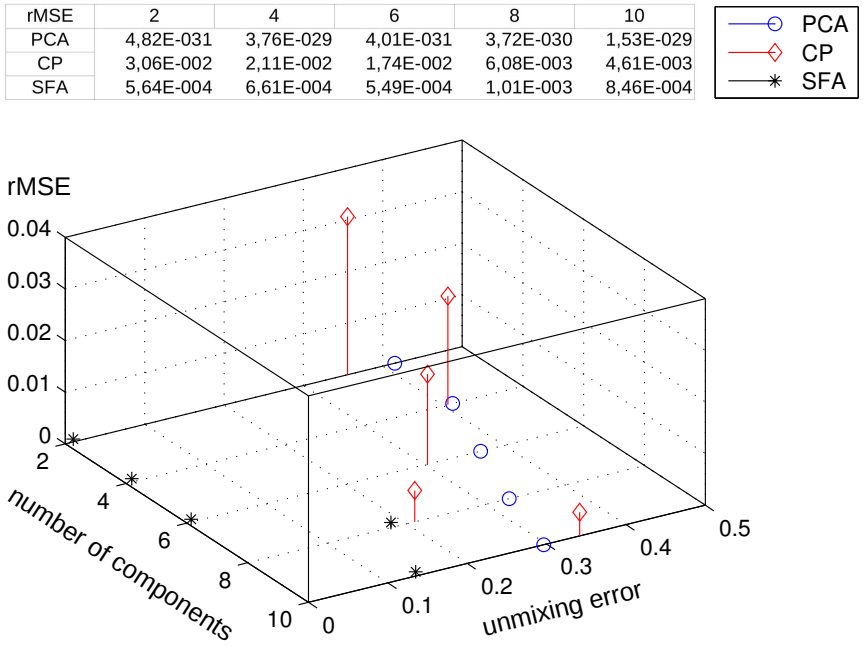


Figure 5.12: Relative fit versus unmixing error of PCA, CP and SFA. Data dimensions were 10 × 50 × 20, modeling the modes space, time and trial. 2, 4, 6, 8 and 10 evoked signals were fitted, respectively. Each, having a maximal shift of 5 samples. The figure suggests that PCA gives best fit, followed by SFA. CP fits the data worst. This is clear from the fact that CP uses least number of degrees of freedom. Nevertheless, low or high fit does not correspond to low or high unmixing error. PCA has worst unmixing performance, CP performs better, while SFA shows best unmixing performance.

freedom as data points yields perfect fit. Hence, fit must not be confused with separation success. This shall be exemplified in the following.

Let PCA, CP and SFA be fit to some virtual evoked data using a data tensor with dimensions 10 × 50 × 20 with modes space, time and trial, respectively. The underlying source dimensionality sis 2, 4, 6, 8 and 10. Each time, the correct number of underlying components was assumed. All evoked source signals have maximal shift of 5 samples and, thus, are dependent. For using PCA, the data tensor is unfolded yielding ordinary twoway matrix channel × time.
The results suggested by Fig. 5.12 show that the fit is always improved from CP to SFA to PCA. However, only SFA yields unmixing error close to zero. In particular, fit merely shows the amount of data variance that is explained. This value is useful as a measure for data compressing. At a particular rMSE level, the method having least parameters compresses the data most efficiently. In contrast, for assessing separation performance, the fit value can be highly misleading. Thus, rMSE is not used here.

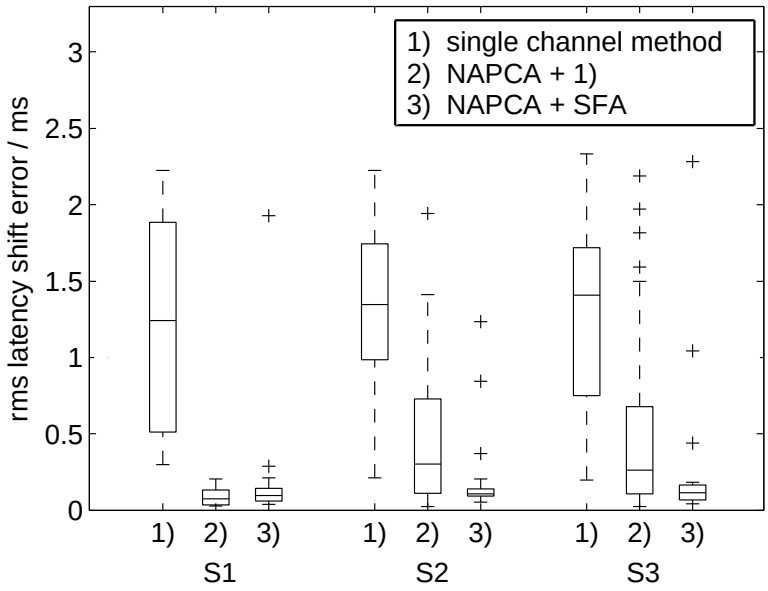


Figure 5.13: Latency estimation performance for maximal latency shifts of 5 ms. For 50 different evoke signal settings with fixed ECD locations S1 - S3, single channel maximum likelihood shift estimates [Jas99] are depicted using raw (1) and NAPCA prefiltered data (2). (3) depicts latency shift estimates obtained by the proposed SFA procedure.

Latency estimation performance

Latency estimation was studied for the case that the latencies were drawn from a uniform distribution $[-\delta, \delta]$ with the maximal possible latency shifts δ chosen equal to 5 ms. Fig. 5.11 shows the results for the latency estimation in one of the 50 signal settings obtained by the proposed procedure and the considered single channel method [Jas99]. The single channel method uses maximum likelihood estimation in a single channel model with fixed signal form and varying amplitudes and latencies. The method was first employed by taking a single channel (optimal for each evoked signal) of the raw data without preprocessing. The channel was chosen such that the resulting latencies best reflected the simulated latency shifts. Note that such a choice is possible only for simulated data. However, it shall be intended to test our method against the best possible choice of the channel taken for the single channel latency estimation procedure. In the results of the single channel method, one observes a correlation between the estimated and the underlying true latencies, but large estimation errors are still present. When the single channel method is applied after NAPCA preprocessing, a significant improvement of the results is observed, which also demonstrates the potential of the NAPCA method. The last row in Fig. 5.11, finally, shows the results of the proposed procedure, which clearly outperform the single channel results obtained with and without preprocessing. Fig. 5.13 summarizes the latency estimation errors from all 50 signal settings considered. The figure shows the root mean square (RMS) errors of the estimated latencies. For each method (and each of the 50 signal settings for the 3 evoked sig-

nals) the RMS errors were determined after applying a common shift to the estimated latencies, thus accounting for the indeterminacy of the absolute latency. The shift was chosen such that the RMS error was minimal. Fig. 5.13 demonstrates that the proposed procedure yields improved latency estimates in most of the cases. Fig. 5.13 further suggests that the latencies for signal 1 are better estimated than those for signal 2 and signal 3. This can be explained by considering the locations of the corresponding ECDs. The optimal single channel sensors for signal 2 and signal 3 (located above these sources) record highly mixed signals, whereas the optimal sensor for signal 1 mainly receives a contribution from the first signal. For larger latencies ($\delta > 15$ ms) a degradation in latency estimation was observed, which may be due to difficulties faced in the estimation procedure caused by local minima. Difficulties also occurred for large latencies with persistent evoked signal cross-correlations.

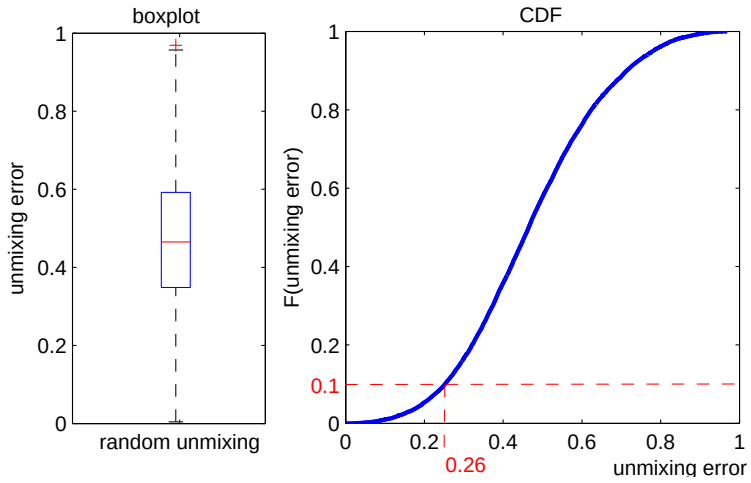


Figure 5.14: Boxplot and cumulative distribution of two-dimensional random unmixing. For 10 % of the results, random unmixing yield an unmixing error equal or below 0.26. A result with less more than 0.26 error to failure. Methods that have unmixing statistics entirely above 0.26 fail, while methods that have unmixing statistics below 0.26 are considered to succeed. They are significantly better than random unmixing at a significance level of 0.1.

5.7 Comparison study

This section shall compare the novel TDSFA method, ICA, bilinear reference methods, CP as well as FDSFA. Virtual experiments as introduced in Sec. 4.2 and the real world audio-visual MEG experiment as detailed in Sec. 4.3.1 are ground truth for comparison.
The actual data experiment has two dimensions, i.e. associated with audition and vision. Hence, all experiments will be carried out in the the two-dimensional setting, being somehow simple. However, time domain support, frequency range, mixing condition, locations of dipoles and more parameters leave high degrees of freedom in simulation.
In order to measure success and failure, two-dimensional random unmixing shall provide a threshold for binary class separation. A boxplot and cumulative distribution of random unmixing are depicted in Fig. 5.14. The statistics and distribution result from 10000 repetitions. For 10 % of the results, random unmixing yields an unmixing error equal or below 0.26. Ad hoc, a result below 0.26 error shall be classified as successful. Methods that have an unmixing level higher 0.26 are classified to having failed. Specifically, methods that yield all boxplot statistics below the threshold are believed to be most useful. The threshold can be interpreted as a significance test. The null hypothesis is that the result is not better than random unmixing. Below 0.26 unmixing error an results is said to be significantly better than random unmixing at 0.1 significance level. Clearly, only an unmixing error close to zero corresponds to true unmixing.
Next to unmixing error statistics, unmixing vector angular deviation statistics are also provided. They shall give more insight into the performance of a method under test, especially in the actual data scenario where the number of experiments is limited to 9 datasets.

5.7.1 Virtual MEG data

To evaluate the performance of TDSFA in contrast to established techniques, two virtual experiments are set up:

- At each repetition of the second experiment, the source signal shapes are generated at random (Eq. 4.8), such that a high variety of signal shapes with low and high-frequency contents occur. The source locations are assumed to be spaced sufficiently apart from each other. In each hemisphere, one source is randomly placed near the auditory cortex. The location setting assures that the associated field maps are not or only partly overlapping. Latency variation, i.e. time shifts, are uniformly distributed from -1 ms to 1 ms. Amplitude variations are in the range of 0.85 to 1. The sampling frequency corresponds to 500 Hz.
- At each repitition of the second experiment, the source signal shapes are generated at random (Eq. 4.8), such that a low variety of signal shapes with mainly low-frequency contents occur. The source locations are assumed be more or less spaced nearby. In each hemisphere, one source is randomly placed near the motor cortex. The location setting assures that the resultant field maps are more or less overlapping. Latency variation, i.e. time shifts, are uniformly distributed from -5 ms to 5 ms. Amplitude variations are in the range of 0.85 to 1. The sampling frequency corresponds to 500 Hz.

100 trials with 100 ms prestimulus duration and 300 ms poststimulus duration are simulated. In total, 100 repetitions are run, each modeling two sources, by placing two ECDs into an isoconducting sphere model. Virtual data was obtained using forward calculation [Sar87]. Furthermore, 50 noise ECDs with different AR processes are randomly placed into the sphere and the corresponding field components are superimposed to the data mixture. 30 dB SNR sensor noise is added. More details are found in virtual experiment setup Sec. 4.2.
In Fig. 5.15 and Fig 5.16 the two experiments are illustrated by showing 4 signal forms and field map distribution pairs that are typical for the first and second experimental settings, respectively. Signal processing chain starts with bandpass filtering between 0.5 Hz and 100 Hz. All trials are detrendet and the offset is removed (refer to Sec. 4.3.1 for more details). Subsequently, NAPCA (Sec. 5.1.2) was applied in order to recover the subspace of evoked dependent components. Finally, the preprocessed data was fed to all decomposition methods under test in sequence.

Fig. 5.17 depicts the results for the first virtual experiment. Next to the unmixing error, angular deviations from the true unmixing vector are provided. This is done to be in line with the actual experiment, where angular deviation shall give more insight as only very few experiments are evaluated. Due to permutation indeterminacy, mapping of unmixing vector and mixing vector was done such that the summed angular deviation is smallest. Fig. 5.17, first row, depicts the larger angular error, while the second row depicts the smaller angular error. The last row depicts the unmixing error. Clearly, a method is said to perform without error, if median and standard

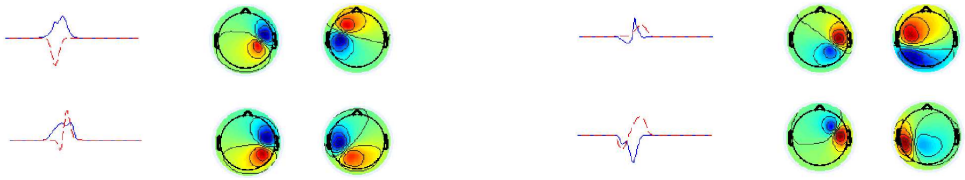


Figure 5.15: Signal shapes and field maps from four runs of the first virtual experiment. Source settings are such that the evoked signal shape show high activity. The field maps are not or only partly overlapping. The signals are dependent with maximal time shift equal 1 ms.

deviation of these performance measures are all equal zero. To further assess performance, 0.26 unmixing error is marked by a dotted line. This threshold shall allow the classification of good and bad methods. For more details see above as well as Fig. 5.14. Median unmixing statistic below 0.26 has good performance on average, while methods with median and standard deviation below 0.26 perform most robust and are considered useful for separation in the context of evoked signals and in the context of the subclass of signals being simulated. Note that the two experiments try to split evoked MEG in roughly two classes in order to allow for a large coverage of signal settings, while gaining insight into different mechanism of the decomposition methods under test.
Specifically, the results in Fig. 5.17 suggest that PCA fails to decompose signals of the first experiment. Only a few outliers yield an unmixing error below 0.26 and PCA is considered not to be suited for decomposition in this context. This is clear as the sources are not constrained to orthogonal mixing and become orthogonal with low probability.
All established ICA methods have an unmixing error statistic with 75 % of the results above 0.26. Hence, ICA is observed not to succeed on the average. The plots suggest that ICA is not robust against dependencies. Angular deviations suggest that some ICA results are recovered almost perfectly. Notably, the whiskers of all boxplots reach zero, validating that ICA is able to recover dependent components in some cases. As discussed in Sec.4.1.3 and 5.3 at length, it is not the level of dependence but the level of ICA dependence reduction that is crucial. However, a relation between median mutual information and unmixing performance exist as shown in Sec. 4.2. Here, it can be observed that ICA fails for most signal configurations, while performing without errors for a small subset of experiments. Notably, the first quartile of the unmixing error distribution is below 0.26. Hence, one in four settings yield an estimate that is only obtained by one in ten trials of random guesses. This suggests that ICA still is worth considering as a reference method or as an initialization method in the dependent signal scenario.
The reference methods and CP yield good average results. Nevertheless, upper quantiles of the unmixing performance distribution suggest that only hpICA, sICA, innoICA and SFA yield very good results. This can be explained by considering the properties of this experiment. The high activity of the signal shape allow the highpass filtering approach and innovations approach to reduce source dependencies yielding improved ICA results. The performance of sICA is explained by the spatial setting of this experiment. All field patterns are not or only partly overlapping and sICA is based on this very assumption.

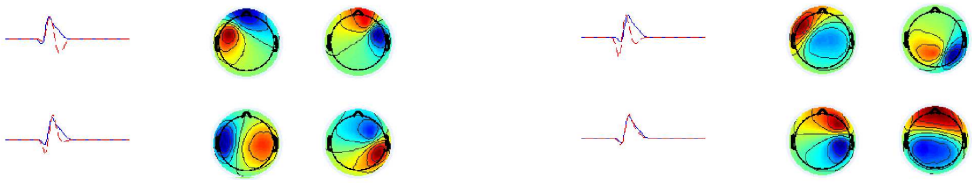


Figure 5.16: Signal shapes and field maps for four runs of the second virtual experiment. Source settings are such that the evoked signal shape possess low activity. The field maps are partly or fully overlapping. The signals are dependent with maximal time shift equal 5 ms.

Notably, the SFA methods TDSFA and FDSFA outperform all methods. Both methods show excellent unmixing performance far below the 0.26 threshold. The associated angular deviations are below 5 degrees for most results.

Fig. 5.18 depicts the results of the second experiment. It confirms that PCA is not suited for source separation in this context. Again, the sources are not constrained to orthogonal mixing and become orthogonal with low probability.
It is worth noting that all ICA methods but SOBI show worse performance than in the first experiment. SOBI performs relatively better but not better than in the first experiment. It still fails on average. Nevertheless, SOBI may be considered valuable for initialization. A consideration of the other established ICA methods is not suggested for signals in this experiment as all results for FastICA, Jade and Infomax are above 0.26 unmixing error. As a consequence, the original source signals were always changed if they were fed to ICA before mixing.
All reference methods and CP show significantly worse results than in the first experiment. This shows that these methods may be more robust to dependencies but sensitive to other source signal assumptions. Spatial ICA fails as field maps are overlapping. CP runs into problems because of temporal shifts. Prefiltering based method seems not to lower dependency enough.
In particular, TDSFA and FDSFA outperform all methods. Both boxplots show excellent performance. The TDSFA associated angular deviations are below 5 degrees for most results. More specifically, TDSFA is observed to yield better results than FDSFA. This may be explained by the fact that FDSFA operate in the frequency domain being more sensitive to transformation errors when the frequency content of the signals are similar. However, FDSFA still has excellent performance - at least for recovering one source.

The SFA methods are at great advantage as the underlying data do follow (almost[1]) a shifted trilinear generative process. Hence, real world experiments are needed to further assess the usefulness of SFA. Actual audio-visual MEG data are considered in the next section.

[1]NAPCA may not work perfectly and can yield data that are not completely corresponding to the shifted factor analysis model.

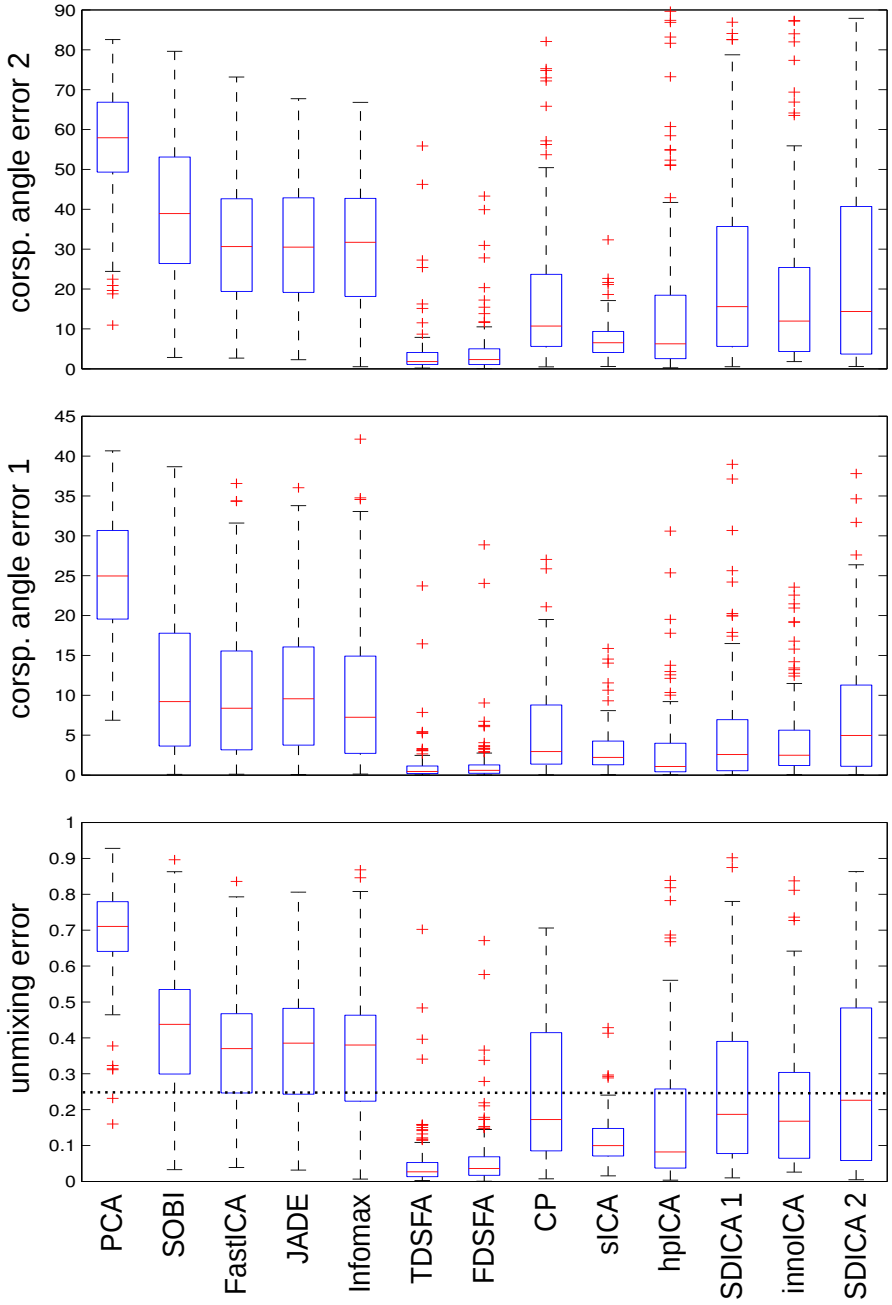


Figure 5.17: Results from the first virtual experiment. PCA and ICA are observed to fail. The prefiltering methods enhance ICA. Furthermore, sICA shows good performance. CP is observed to yield reasonable results. However, TDSFA and FDSFA outperform all methods being tested and show excellent performance.

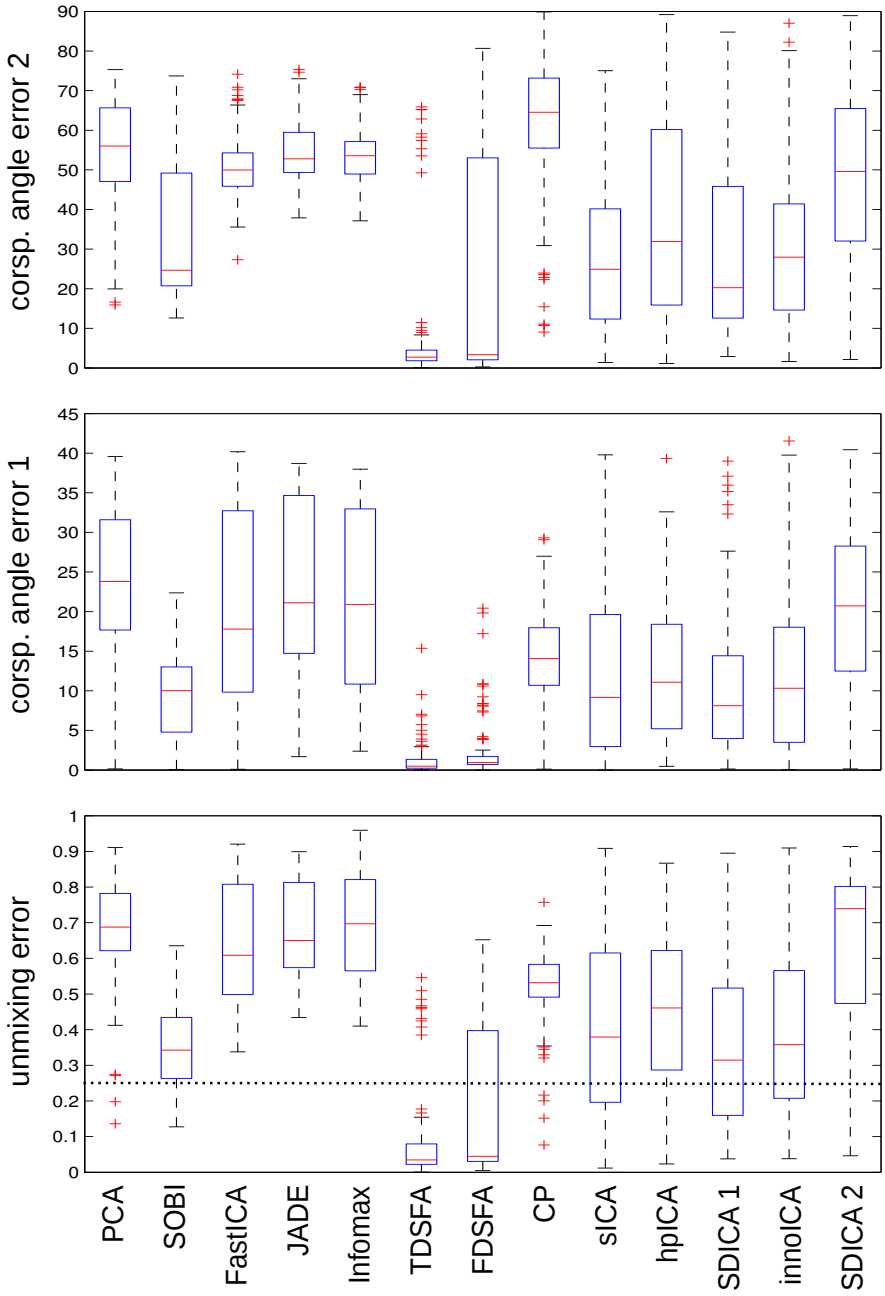


Figure 5.18: Results from the second virtual experiment. PCA and ICA are observed to fail. All reference methods show poor performance. TDSFA and FDSFA outperform all methods under test. Indeed, both have excellent performance, while TDSFA is at slight advantage in this setting.

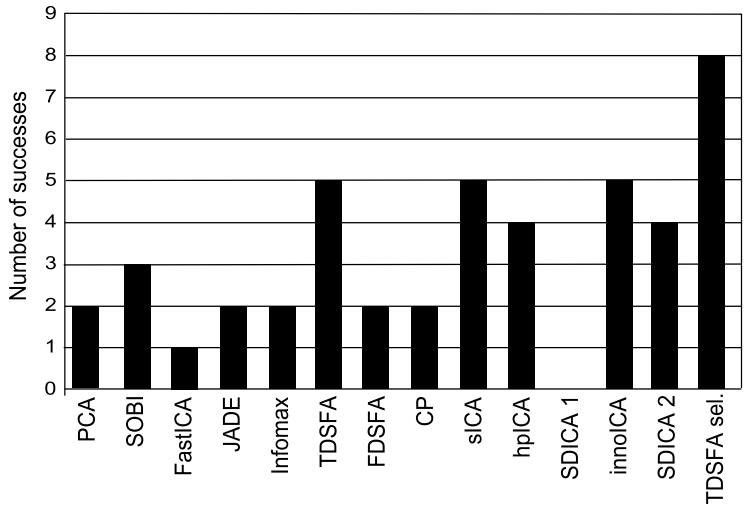


Figure 5.19: Number of successes for actual audio-visual MEG data decomposition; associated unmixing error is below 0.26. InnoICA, hpICA, SDICA2 as well as sICA show good performance. The latter is at advantage as sources are spaced far apart. TDSFA shows better results than CP and FDSFA and equal results with sICA and innoICA. Specifically, TDSFA is more suited than ICA as well as CP for these evoked data. The last row suggests that TDSFA has even higher potential in performing. 10 trials are selected by visual inspection prior to running TDSFA yielding all but one experiment with low unmixing error.

5.7.2 Actual MEG data

Actual audio-visual evoked MEG data are used to further assess performance in a real word scenario using the 9 MEG datasets as presented in section 4.3.1. All details on preprocessing as well as experimental setup are found in section 4.3.1. They include bandpass filtering, detrending and offset and bad trial removal. NAPCA was used to construct a mixing matrix from unisensory auditory and visual data. The bisensory audio-visual stimulated data is projected onto the columnspace of that mixing matrix. This procedure allows knowing the ground truth in an actual data scenario - assuming locations of activity to be stationary. Finally, the preprocessed data was fed to the method under test.
In Fig.4.14, ICA was applied to the 9 datasets and evaluated in detail by using a novel single-trial performance illustration technique. ICA was observed to have failed in separation of seven datasets as suggested by Fig. 4.14. The same figure shall be constructed using TDSFA for comparing the results. In particular, starting from the ICA, TDSFA performs as depicted in Fig. 5.21. The single-trial plots show that five recordings are separated. More than double the number of datasets have unmixing errors below 0.26 in comparison with ICA. Still, four datasets are not decomposed correctly. Choosing only the 10 best trials (selected by visual inspection) yield a separation with associated unmixing error below 0.26 for 8 datasets! This suggests that SFA has high potential in the case that trials follow the model assumed. In turn, not all trials seem to follow the SFA model, which pose difficulties for this technique. Specifically, trials that occurred in

the second half of the experiment as well as a first few are observed to lead SFA to worse results than deleting those trials. This may be explained with the concentration level of the subject. At the beginning and at the end of the experiment, the concentration is less accurate yielding data that is not compliant with the model assumptions.
Fig. 5.19 shows the frequency of separation below 0.26 unmixing error for each method. The results clearly suggest that SFA is a candidate method for the separation of actual domain MEG data. The last column of 5.19 shows the excellent potential of SFA if only 'good' trials are considered. In particular, SFA seems to be more successful than CP and ICA. In Fig. 5.20, TDSFA is further compared to all methods in more detail. Very similar to the virtual experiment, the boxplot statistics are provided in Fig. 5.20. This time, angular deviation shall provide more information as only 9 datasets are available. Again, 0.26 unmixing error shall be used as threshold in order to assess performance. In particular, Fig. 5.20 suggests that median unmixing performance below the threshold is attained by innoICA, spatial ICA as well as TDSFA. The results of spatial ICA are expected as field maps are only partly overlapping. Notably, SDICA2 and hpICA still perform relatively well. However, it is not well understood by the author why FDSFA performs rather poorly. Fine tuning with different initialization, termination and number of iterations as well as constraints such as orthogonality and non negativity of the amplitudes mode has been considered. For the results depicted, ICA is used as initialization and no constraints are imposed, such as in the TDSFA setting. Notably, all methods are largely outperformed by utilizing 'cheated' TDSFA. Specifically, choosing trials by visual inspection (ten trials from the first 10 minutes of the experiment) yields results as suggested by the last column in Fig. 5.20. Eight of the nine datasets are separated! Clearly, this underlines the high potential of SFA and shows that the model assumptions are not always met in all trials.
Consequently, a method that deletes non-compliant trials may be helpful. Notably, this technique can have a positive effect for CP and ICA as one may argue that only some trials are dependent and only some trials are shifted. However, if the trials are assumed to be shifted and dependent, SFA combined with sophisticated trial selection may have higher potential to benefit from a trial selection technique. If less trials need to be rejected, SFA may be more robust than ICA and CP as it covers presumably more compliant trials. This approach and its consequences are beyond the scope of this work and need to be addressed in the future.

5.8 Discussion

This section discusses alternatives to ICA. It was shown that the use of ICA results based on mutual information for subspace or separation analysis is questionable. Indeed, ICA optimizes some feature of independence. At a global minimum no more reduction of this feature is possible under linear transformation. However, the unmixing performance may be far from optimal.
This can be compared to least squares fitting in CP and SFA. A good fit only says that the model explains variance. Yet, it was designed to do so! Arguing along the same line, ICA is designed to yield independent components. Hence, for both techniques, the optimized value cannot be linked to separation success. Clearly, a perfect fit is always obtained using a very large number of components; independent components are always obtained when the dependent sources can be made independent under linear transformation. A good fit as well as a reasonable level of independence cannot tell separation performance.
As a consequence, virtual and actual experiments were designed to evaluate the separation performance and to search for alternatives. Shifted factor analysis was shown to yield better results than ICA or classic CP. SFA introduces more degrees of freedom to the CP model, while retaining the favorable property of essential uniqueness. The data are assumed to follow this model and must be trilinear up to one mode that is allowed to be shifted. Raw MEG data may not follow the SFA model as non-evoked sources may be far from being shifted trilinear.
As a consequence, the use of NAPCA is highly advocated. NAPCA aims at finding structure that corresponds to the SFA model, i.e. that contain the evoked sources. These are assumed to be trilinear with shifts along the temporal mode and, hence, the combination of NAPCA and SFA is assumed to be suited for decomposition of evoked MEG data.
The novel TDSFA method approaches SFA in the time domain. Taylor series are employed and truncated in practical use. Derivatives were approximated numerically. Nevertheless, TDSFA is robust to the design choices. Low and high numbers of Taylor components both give good performance, while using simple central differences for approximating the derivatives worked out. The comparison study showed that for both virtual and actual data experiment, TDSFA yields very good performance without fine tuning its parameters, such as number of Taylor components or numerical approximation of differentiation. In contrast, ICA as well as CP are observed to fail in more cases.
In particular, this chapter has shown high potential for the combination of NAPCA and TDSFA and showed problems of ICA as well as CP due to dependencies as well as shifts. ICA performs worse if no shifts occur and CP is the right model to choose. Shifts lower dependencies. Still, CP and ICA are observed to fail and SFA should be used. For very high shifts, signals are made independent and ICA should be used. However, in practice, small shifts are likely and SFA is assumed to be most appropriate. Indeed, actual data suggest that TDSFA is suited for real word data. ICA and its generalizations, such as ICA on innovations, often yield good initial solutions. They were considered as preprocessing techniques prior to the application of TDSFA.

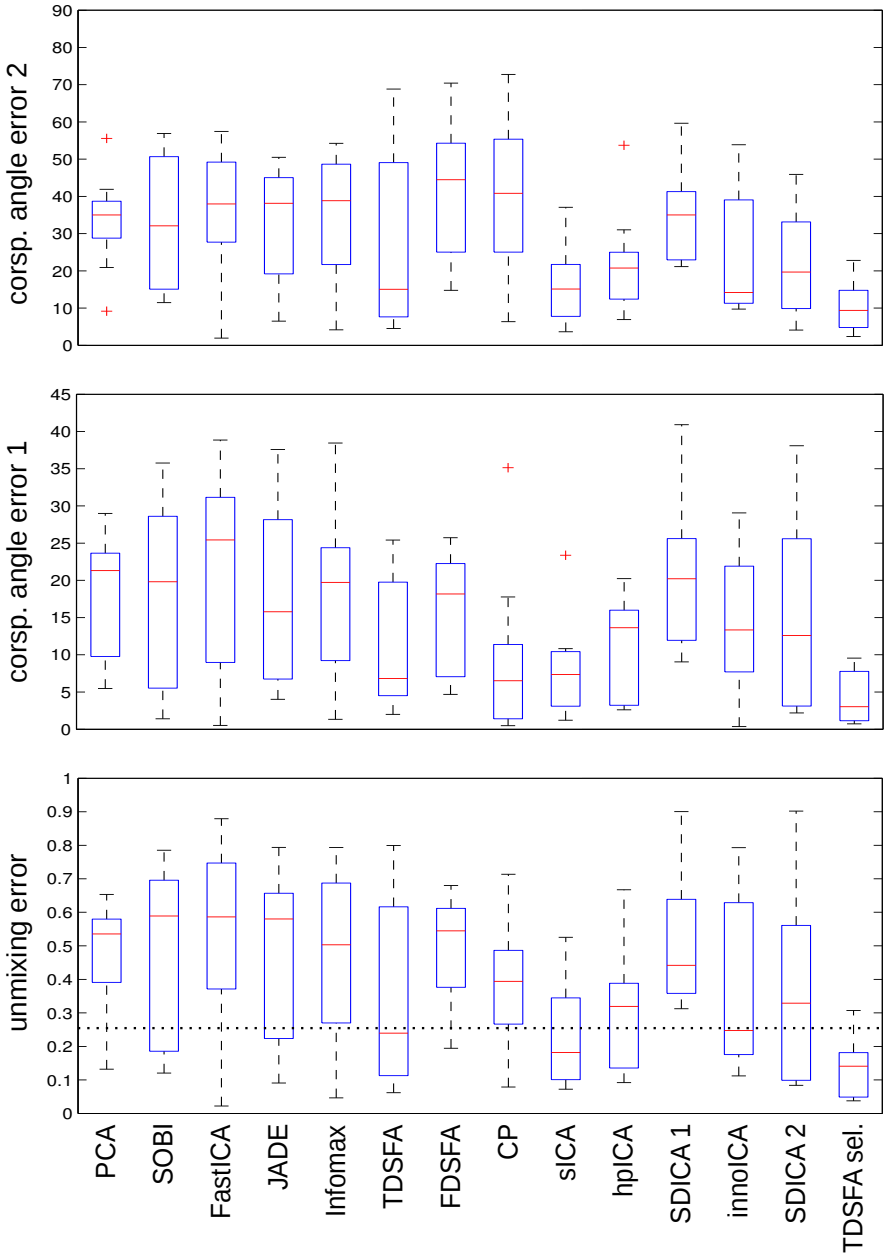


Figure 5.20: Decomposition results from actual audio-visual MEG data. PCA and ICA are observed to fail in most cases. Median unmixing performance below 0.26 unmixing error is reached by innoICA, sICA and TDSFA. SFA outperforms ICA and CP, which underlines the potential of a flexible model that is still essentially unique. The last row is cheated in the sense that only 10 trials are selected by visual inspection prior to running TDSFA. However, the boxplot suggests that TDSFA has even higher potential in performing; all but one dataset is decomposed with low unmixing error by TDSFA.

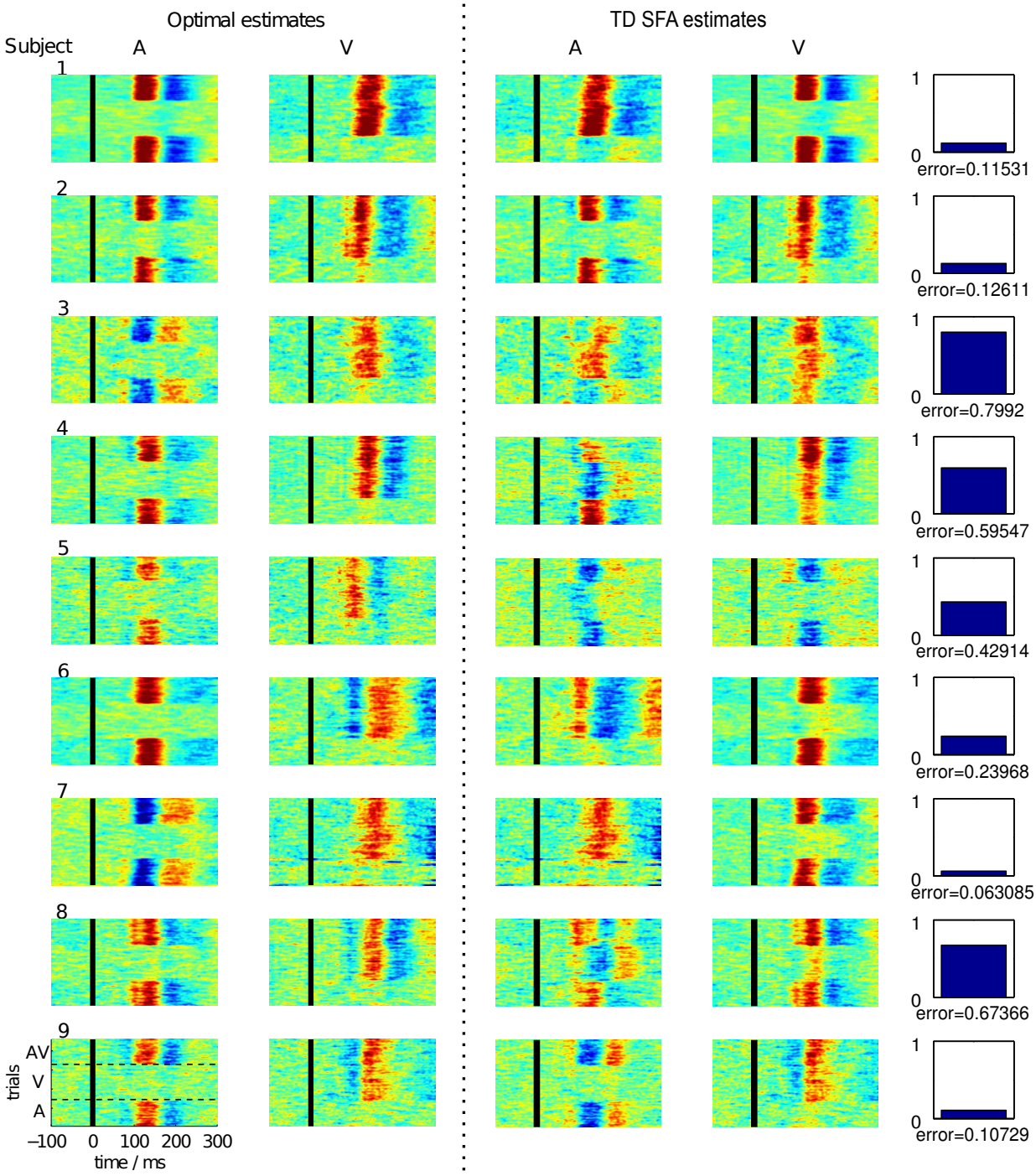


Figure 5.21: Results of applying TDSFA to the data from the audio-visual experiment (AV data). The TDSFA unmixing vectors are applied to AV, A and V data, respectively. In the rightmost column the unmixing error is provided. The obtained and stacked single-trial plots show that five recordings are separated by TDSFA. In comparison with Fig. 4.14, the TDSFA approach clearly outperforms ICA.

6 Summary

This thesis issued the decomposition of evoked MEG data into the data underlying dependent source signals employing a linear generative model assumption. Under the term blind source separation, the framework of independent component analysis has gained widespread popularity. ICA is a generalization of PCA as it is able to disentangle any regular mixing matrix, whereas PCA is limited to orthogonal mixing. Separation of non-orthogonal mixing is possible assuming independence among and non-Gaussianity of the sources. If these assumptions are met, ICA is essentially unique and yields the underlying components, i.e. brain signals of interest as well as brain or non-brain signals that are not of interest.

MEG data from stimulation experiments build a subset of general MEG data. One feature of evoked data is its repetitive nature - evoked components are structured in trials. In each trial, one or more signals may be evoked at one or more brain locations. Using evoked data, ICA often recovers at least one component that possess a dipolar field map as well as typical structure in the time or frequency domain. Often, such a component is said to be physiologically plausible. Indeed, it is tempting to refer to such a component as *the* evoked signal. However, more than one source may be evoked by one or more stimuli. As a consequence, ICA's assumption of independence may no longer hold true.

Whether plausible components give evidence to true components was not addressed in the context of evoked experiments. In particular, an ICA robustness analysis of dependence was missing. Consequently, this thesis posed and discussed questions such as 'Can ICA separate dependent components?' or 'How much dependence is allowed?'. It was demonstrated that highly dependent signals can be separated by ICA, if, under linear transformation, the original sources cannot be made less dependent. Indeed, ICA will reduce dependence introduced by the mixing process. As it cannot further lower dependency, the original dependent signals are perfectly separated. On the other hand, if, under linear transformation, dependent sources are changed by an ICA, the unmixing matrix $\mathbf{W}$ will differ from identity. Consequently, if these dependent source signals are mixed, ICA will also act on the sources yielding the unmixing matrix $\mathbf{W}$ again . Indeed, ICA lowers the dependency among the estimates that were introduced due to linear mixing but, furthermore, lowers the dependency of the original signals as much as possible under linear transformation. In these cases ICA is deemed to fail. The question is, 'Can ICA separate typical evoked dependent components?'.

To assess the performance and usefulness of ICA in the evoked dependent signal scenario, a virtual MEG experiment as well as an actual MEG data experiment was designed. Using virtual data, the results showed that median unmixing error closely follows mutual information of the original signals. Furthermore, for closely spaced location of the sources, ICA components showed physiologically plausible dipolar structure. Such components are often believed to be correctly separated but were shown still to be a mixture - ICA did not succeed. For nine audio-

visual evoked data sets with dependent sources, ICA was observed to fail in seven cases. This means that ICA is likely to provide only some vector of an evoked signal subspace. This recovered component is not an underlying source but a mixture of evoked sources in that signal subspace. Thus, the first challenge towards a technique that can separate independent and dependent sources is subspace analysis. This step can be considered as a preprocessing step. It is a useful step as the dimensionality of the recorded data is often drastically reduced by considering only the subspace of evoked signals. Furthermore, noise can be reduced and the structure underlying the data can be assumed to be more homogeneous. The data consists mainly of evoked contribution, which facilitates analysis.
Cardoso et al. conjectured that ICA can separate data with underlying dependent and independent sources into independent sources and subspaces containing mixed dependent sources. A grouping algorithm based on residual dependencies was proposed to determine these subspaces. Nevertheless, it was demonstrated that using a binned mutual information ICA lowers binned mutual information of the sources. Hence, ICA fails but the user may not be aware of it as the estimates may only show low dependence. It was concluded that mutual information among ICA estimates is a questionable quantity to use in order to group components to subspaces or to infer the success of ICA. As a consequence, noise adjusted principal component analysis was considered. NAPCA makes use of pre- and poststimulation information, using simple and robust covariance statistics. After whitening the noise space, a second PCA of sphered noise plus transformed evoked signal yield eigenvectors that point into the direction of evoked activity. By this, NAPCA robustly recovers the evoked subspace. The use of NAPCA was highly advocated due to its robustness and ease of use. For single evoked signal, NAPCA can serve as a separation and denoising tool yielding highly accurate estimates. Hence, in all experiments, NAPCA was run as a preprocessing method with excellent performance on virtual and actual MEG data. Notably, finding the subspace of evoked signals is of utmost interest for specific evoked data models, such as CP. Which in turn is of high interest as CP is essentially unique without assuming independence. In contrast, the bilinear model is essentially non-unique.
Decomposition methods that aim at the data underlying sources have to impose assumptions. Specifically, all methods that use a bilinear mixing model have to assume some objective function in order to render the solution essentially unique. It is questionable whether a new objective is more robust than assuming independence for evoked MEG.
Stimulation experiments are structured, a general property, which may be further exploited. E.g., CP aims at using this structure assuming trilinear components with modes space, time as well as trial. Nevertheless, for a successful CP decomposition to work, the recorded data must follow a trilinear generative process. As shifts are common to actual data, the model is per se incorrect. In line with the assessment of ICA to the violation of the independence assumption, CP needed to be assessed for its robustness to shifts. Again, in the context of evoked MEG, this was not done before. At some level, shifts must destroy the separation performance of CP. It was assessed in the context of shifts using different signal shapes as well as temporal support of the virtual evoked signals. Depending on the signal shape, already small shifts of a single sample was shown to destroy a CP decomposition. Only for signals with low frequency content CP did tolerate shifts up to 3 samples, in the case that temporal support is about 200 samples. These findings suggested that CP can be highly sensitive to model violation by shifts in one mode. Clearly, in actual data,

List of contributions	Section	Figure	Publication
Virtual simulation tool for evoked dependent sources	4.2	4.4	[Koh08a, Koh09]
Robustness assessment of ICA to dependence	4.2.5	4.7	[Koh08b, Koh09]
The nearby dipole effect	4.2.6	4.9	[San10, Koh09]
Discussion on mutual information of ICA results	5.3	5.4	-
Proposition of using NAPCA for subspace analysis	5.1.2	5.2	[Koh10b]
Robustness assessment of CP to shifts	5.4.4	5.7	-
Time domain shifted factor analysis (TDSFA)	5.6	5.9	[Koh10a]
Audio-visual paradigm yielding dependent sources	4.3	4.10	[Koh10b]
Single-trial performance illustration technique	4.3.4	4.11	-
Comparison study based on actual and virtual data	5.7	5.21	-

shifts may not be integer spaced and can still have impact on the CP performance.

Harshman and Hong proposed a novel method including shifts. Shifted factor analysis was shown to be essentially unique by Morup. He further presented the first method in the frequency domain, which is capable of estimating non-integer shifts. Field et al. noted that when applying CP to evoked shifted data, components were doubled with the same spatial profile but with differentiated temporal signatures. Based on this remarkable observation and the work of Morup, this work proposed to approximate the shifted factor analysis model by using a Taylor series based expansion. It turned out that for infinite many Taylor components, SFA equals the CP model that has infinite many components. Truncating the Taylor series gave a robust algorithm that largely outperformed ICA, CP and other reference techniques. In contrast to FDSFA, a linear algorithm is obtained when using first order Taylor approximation. For higher order approximation, say a 10th order approximation, TDSFA was shown to perform better on the actual audio-visual MEG data. ICA and CP only succeed in one dataset, while TDSFA succeeded in 5 of 9 recordings. Nevertheless, bad trials can lower the performance. In particular, it was demonstrated that manual selection of 10 trials by visual inspection yielded excellent performance of TDSFA with 8 of 9 successful decompositions.

It is worth further investigating shifted factor analysis. One line of future work is automation of trial selection. Another line is the investigation of the number of degrees of freedom, e.g. for allowing signal shape variations from trial to trial, or other variations. This, however has to be limited in order to keep the favorable essential uniqueness property.
Another line of investigation is to reformulate the multi-stage SFA algorithm as a single-stage algorithm. Having all stages sharing information should improve the results of each stage. Lastly, a method is only as good as its numerical implementation. As well as experimental stimulation paradigm design and care in recording, the choice of preprocessing, numerical optimization and its robustness are crucial. Efficient and robust numerical techniques are still to be assessed and designed in order to profit from SFA to the full in practice.

Bibliography

[Abr03] Abrard, F. and Deville, Y. 2003. Blind separation of dependent sources using the time-frequency ratio of mixtures approach. *Proc. ISSPA'03*, 2:81–84.

[Ahl92] Ahlfors, S.P., Ilmoniemi, R.J. and Hämäläinen, M.S. 1992. Estimates of visually evoked cortical currents. *Electroencephalography and clinical neurophysiology*, 82(3):225–236.

[Ain03] Aine, CJ and Stephen, JM. 2003. MEG Studies of Visual Processing. Chapter in *The cognitive electrophysiology of mind and brain*, 93–142, Academic Press, Elsevier.

[Ama96] Amari, S., Cichocki, A. and Yang, H.H. 1996. A new learning algorithm for blind signal separation. *Proc. NIPS'96*, 757–763.

[Ama98] Amari, S.I. 1998. Natural gradient works efficiently in learning. *Neural computation*, 10(2):251–276.

[Ame94] American EEG society. 1994. American EEG Society: Guidelines in electroencephalography, evoked potentials, and polysomnography. *Clinical neurophysiology*, 11:1–147.

[And00] Andersson, C.A. and Bro, R. 2000. The N-way toolbox for matlab. *Chemometrics and intelligent laboratory systems*, 52(1):1–4.

[Bac02] Bach, F.R. and Jordan, M.I. 2002. Tree-dependent component analysis. *Proc. UAI'02*, 18:50–61.

[Bac03] Bach, F.R. and Jordan, M.I. 2003. Beyond independent components: trees and clusters. *The journal of machine learning research*, 4:1205–1233.

[Bai01] Baillet, S., Mosher, J.C. and Leahy, R.M. 2001. Electromagnetic brain mapping. *IEEE signal processing magazine*, 18(6):14–30.

[Bec08] Becker, R., Ritter, P. and Villringer, A. 2008. Influence of ongoing alpha rhythm on the visual evoked potential. *Neuroimage*, 39(2):707–716.

[Bel95] Bell, A.J. and Sejnowski, T.J. 1995. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6):1129–1159.

[Bel97] Belouchrani, A., Abed-Meraim, K., Cardoso, J.F. and Moulines, E. 1997. A blind source separation technique using second-order statistics. *IEEE transactions on signal processing*, 45(2):434–444.

[Ber29] Berger, H. 1929. Über das Elektrenkephalogramm des Menschen. *European archives of psychiatry and clinical neuroscience*, 87(1):527–570.

[Bla04] Blankertz, B., Müller, K.R., Curio, G., Vaughan, T.M., Schalk, G., Wolpaw, J.R., Schlogl, A., Neuper, C., Pfurtscheller, G., Hinterberger, T. and others. 2004. The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials. *IEEE transactions on biomedical engineering*, 51(6):1044–1051.

[Bou03] Bounkong, S., Toch, B., Saad D. and Lowe, D. 2003. ICA for watermarking digital images. *The journal of machine learning research*, 4:1471–1498.

[Bro97] Bro, R. and Jong, S.D. 1997. A fast non-negativity-constrained least squares algorithm. *Journal of chemometrics*, 11(5):393–401.

[Bro97a] Bro, R. 1997. PARAFAC. Tutorial and applications. *Chemometrics and intelligent laboratory systems*, 38(2):149–171.

[Bro97b] Bro, R. and Andersson, C.A. 1998. Improving the speed of multiway algorithms: Part II: Compression. *Chemometrics and intelligent laboratory systems* 42(1-2):105–113.

[Bro98] Bro, R. 1998. *Multi-way analysis in the food industry: models, algorithms, and applications*, PHD thesis, Københavns Universitet.

[Bro03] Bro, R. and Kiers, H.A.L. 2003. A new efficient method for determining the number of components in PARAFAC models. *Journal of chemometrics*, 17(5):274–286.

[Bur05] Burghoff, M., Link, A., Salajegheh, A., Elster, C., Poeppel, D. and Trahms, L. 2005. A template-free approach for determining the latency of single events of auditory evoked M100. *Physics in medicine and biology*, 50: N43-N48.

[Cae03] Caesar, K., Gold, L. and Lauritzen, M. 2003. Context sensitivity of activity-dependent increases in cerebral blood flow. *Proceedings of the national academy of sciences of the united states of america*, 100(7):3550–3552.

[Cai06] Caiafa, C.F. and Proto, A.N. 2006. Separation of statistically dependent sources using an L2-distance non-Gaussian measure. *Signal Processing*, 86(11):3404–3420.

[Cal09] Calhoun, V.D., Liu, J. and Adali, T. 2009. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage*, 45(1):S163–S172.

[Car93] Cardoso, J.F. andSouloumiac, A. 1993. Blind beamforming for non-Gaussian signals. *Radar and signal processing, Proceedings-F IEE*. 140(6):362–370.

[Car96a] Cardoso, J.-F. and Souloumiac, A. 1996. Jacobi angles for simultaneous diagonalization. *SIAM Journal on matrix analysis and applications*, 17(1):161–164.

[Car96b] Cardoso, J.F. and Laheld, B.H. 1996. Equivariant adaptive source separation. *IEEE transactions on signal processing*, 44(12):3017–3030.

[Car98a] Cardoso, J.F. 1998. Blind signal separation: statistical principles. *Proceedings of the IEEE*, 86(10):2009–2025.

[Car98b] Cardoso, J.F. 1998. Multidimensional independent component analysis. *Proc. ICASSP'98*, 4:1941–1944.

[Car99] Cardoso, J.F. 1999. High-order contrasts for independent component analysis. *Neural computation*, 11(1):157–192.

[Car01] Cardoso, J.F. 2001. The three easy routes to independent component analysis; contrasts and geometry. *Proc. ICA'01*, 3:1–6.

[Car03] Cardoso, J.F. 2003. Dependence, correlation and Gaussianity in independent component analysis. *Journal of machine learning research*, 4:1177–1203.

[Car70] Carrol, J.D., Chang, J.J. 1970. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35(3):283–319.

[Cha84] Chatrian, G.E., Goff, W.R., Picton, T.W., Coats, A.C., Celesia, G.G. and Cracco, J.B. 1984. American electroencephalographic society guidelines for clinical evoked potential studies. *Clinical neurophysiology*, 1(1):3–53.

[Cho02] Choudrey, R.A. 2002. *Variational methods for Bayesian independent component analysis*, PhD thesis, University of Oxford.

[Cic02] Cichocki, A. and Amari, S. 2002. *Adaptive blind signal and image processing*, Chicester, Wiley.

[Cic03] Cichocki, A. and Georgiev, P. 2003. Blind source separation algorithms with matrix constraints. *IEICE Transactions on information and systems, special session on independent component analysis and blind source separation*, 86:522–531.

[Cic08] Cichocki, A., Zdunek, R. and Amari, S. 2008. Nonnegative matrix and tensor factorization. *IEEE signal processing magazine*, 25(1):142–145.

[Coh68] Cohen, D. 1968. Magnetoencephalography: evidence of magnetic fields produced by alpha- rhythm currents. *Science*, 161(3843):784–786.

[Com94] Comon, P. 1994. Independent component analysis, a new concept? *Signal processing*, 36(3):287–314.

[Com09] Comon, P., Luciani, X. and De Almeida, A.L.F. 2009. Tensor decompositions, alternating least squares and other tales. *Journal of chemometrics* 23(7-8):393–405.

[Cov91] Cover, T.M. and Thomas, J.A. 1991. *Elements of information theory*, New York, Wiley.

[Cve07] Cvejic, N., Bull, D. and Canagarajah, N. 2007. Improving fusion of surveillance images in sensor networks using independent component analysis *IEEE transactions on consumer electronics*, 53(3):1029–1035.

[Dal93] Dale, A.M. and Sereno, M.I. 1993. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *Journal of cognitive neuroscience*, 5(2):162–176.

[Dar53] Darmois, G. 1953. Analyse generale des liaisons stochastiques. *Review of the international statistical institute*, 21(1/2):2–8.

[DaS91] Da Silva, F.H. and Spekreijse, H. 1991. Localization of brain sources of visually evoked responses: using single and multiple dipoles. An overview of different approaches. *Electroencephalography and clinical neurophysiology supplement*, 42:38–46.

[Daw54] Dawson, G.D. 1954. A summation technique for detecting of small evoked potentials. *Electroencephalography & clinical neurophysiology*, 6(1):65–84.

[DeL97] De Lathauwer, L. 1997. *Signal processing based on multilinear algebra*. PhD thesis, Leuven, Katholieke Universiteit Leuven.

[Del04] Delorme, A. and Makeig, S. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21.

[DeS08] De Silva, V. and Lim, L.H. 2008. Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on matrix analysis and applications* 30(3):1084–1127.

[Fan07] Fang, Y. and Takaya, K. 2007. An application of ICA to DS-CDMA detection.*Proc. CCECE'07*, 609–612.

[Fie91] Field, A.S. and Graupe, D. 1991. Topographic component (parallel factor) analysis of multichannel evoked potentials: practical issues in trilinear spatiotemporal decomposition.. *Brain topography*, 3(4):407–423.

[Fra03] Frackowiak, R.S.J., Friston, K.J., Frith, C.D., Dolan, R.J., Price, C.J., Zeki, S., Ashburner, J.T. and Penny, W.D. 2004. *Human Brain Function*, San Diego, Elsevier Science.

[Gha10] Ghaemi, D., Kohl, F. and Orglmeister, R. 2010. Classifying ICA components of evoked MEG data. *Proc. BMT*, 55:302–305.

[Gil81] Gill, P.E., Murray, W., and Wright, M.H. 1981. *Practical optimization*. London, Academic Press.

[Gol96] Golub, G.H. and Van Loan, C.F. 1996. *Matrix computations*. Baltimore, Johns Hopkins University Press.

[Gra18] Gray, H. 1918. *Anatomy of the Human Body.* Philadelphia, Lea & Febiger, www.bartleby.com/107/.

[Gut07] Gutch, H.W. and Theis, F.J. 2007. Independent subspace analysis is unique, given irreducibility. *Proc. ICA'07*, lecture notes computer in science, 4666:49–56.

[Hai64] Haider, M., Spong, P. and Lindsley, D.B. 1964. Attention, vigilance, and cortical evoked-potentials in humans. *Science*, 145:180–182.

[Har05] Hari, R. 2005. Magnetoencephalography in clinical neurophysiological assessment of human cortical functions. in *Electroencephalography: basic principles, clinical applications, and related fields*, 1165–1197, Philadelphia, Lippincott Williams & Wilkins.

[Har70] Harshman, R.A. 1970. Foundation of the PARAFAC procedure: models and conditions for an explanatory multi-modal factor analysis. *UCLA working papers in phonetics*, 16:1–84.

[Har72] Harshman, R.A. 1972. Determination and proof of minimum uniqueness conditions for PARAFAC1. *UCLA working papers in phonetics*, 22:111–117.

[Har03] Harshman, R.A., Hong, S. and Lundy, M.E. 2003. Shifted factor analysis–Part I: models and properties. *Journal of chemometrics*, 17(7):363–378.

[Hau02] Haueisen, J. and Tuch, D.S. and Ramon, C. and Schimpf, PH and Wedeen, VJ and George, JS and Belliveau, JW. 2002. The influence of brain tissue anisotropy on human EEG and MEG. *Neuroimage*, 15(1):159–166.

[Häm93] Hämäläinen, M., Hari, R., Ilmonieni, R.J., Knuutila, J. and Lounasmaa, O.V. 1993. Magnetoencephalography-theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern physics*, 65(2):413–497.

[Hit27] Hitchcock, F. L. 1927. The expression of a tensor or a polyadic as a sum of products. *Mathematics and physics*, 6:164–189.

[Hon03a] Hong, S. and Harshman, R.A. 2003. Shifted factor analysis–part II: algorithms. *Journal of chemometrics*, 17:379–388.

[Hon03b] Hong, S. and Harshman, R.A. 2003. Shifted factor analysis-Part III: N-way generalization and application. *Journal of chemometrics*, 17:389–399.

[Hyv97] Hyvärinen, A. and Oja, E. 1997. A fast fixed-point algorithm for independent component analysis. *Neural computation*, 9:1483–1492.

[Hyv98] Hyvärinen, A., Independent component analysis for time-dependent stochastic processes. *Proc. ICANN'98*, 135–140.

[Hyv00] Hyvärinen, A. and Hoyer, P. 2000. Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural computation*, 12(7):1705-1720.

[Hyv01a] Hyvärinen, A., Karhunen, J. and Oja, E. 2001. *Independent component analysis*, New York, Wiley.

[Hyv01b] Hyvärinen, A., Hoyer,P.O., Inki, M.O. 2001. Topographic independent component analysis. *Neural computation*, 13(7):1527–1558.

[Hyv04] Hyvärinen, A. and Hurri, J. 2004. Blind separation of sources that have spatiotemporal variance dependencies. *Signal processing*, 84(2):247–254.

[Hyv06] Hyvärinen, A. and Köster, U. 2006. FastISA: a fast fixed-point algorithm for independent subspace analysis. *Proc. ESANN'06*, 371–376.

[Jam05] James, C.J. and Hesse, C.W. 2005. Independent component analysis for biomedical signals. *Physiological measurement*, 26(1):R15–39.

[Jas99] Jaskowski, P. and Verleger, R. 1999. Amplitudes and latencies of single-trial ERPs estimated by a maximum likelihood method. *IEEE Transactions on biomedical engineering*, 46(8):987–993.

[Jöb77] Jöbsis, F.F. 1977. Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science*, 198(4323):1264–1267.

[Jun01] Jung, T.P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E. and Sejnowski T.J. 2001. Analysis and visualization of single-trial event-related potentials. *Human brain mapping*, 14: 166–185.

[Kad99] Kado, H., Higuchi, M., Shimogawara, M., Haruta, Y., Adachi, Y., Kawai, J., Ogata, H. and Uehara, G. 1999. Magnetoencephalogram systems developed at KIT. *IEEE transactions on applied superconductivity*, 9(2):4057–4062.

[Kie91] Kiers, H.A.L. 1991. Hierarchical relation among three-way methods.*Psychometrika*, 56(3):449–470.

[Kie00] Kiers, H.A.L. 2000. Towards a standardized notation and terminology in multiway analysis. *Journal of chemometrics*, 14:105–122.

[Kim03] Kim, S.G. 2003. Progress in understanding functional imaging signals. *Proceedings of the national academy of sciences*, 100(7):3550–3552.

[Kis99] Kisley, M.A. and Gerstein, G.L. 1999. Trial-to-trial variability and state-dependent modulation of auditory-evoked responses in cortex. *Neuroscience*, 19(23):10451–10460.

[Kli06] Klimesch, W., Hanslmayr, S., Sauseng, P. and Gruber, W.R. 2006. Distinguishing the evoked response from phase reset: a comment to Mäkinen et al. *Neuroimage*, 29(3):808–811.

[Knu06] Knuth, K.H., Shah, A.S., Truccolo, W.A., Ding, M., Bressler, S.L. and Schroeder, C.E. 2006. Differentially variable component analysis: identifying multiple evoked components using trial-to-trial variability. *Neurophysiology*, 95(5):3257–3276.

[Koh08a] Kohl, F., Wübbeler, G., Sander, T., Trahms, L., Kolossa, D., Orglmeister, R., Elster, C. and Bär, M. 2008. Performance of ICA for dependent sources using synthetic stimulus evoked MEG data. *Proc. DGBMT-Workshop Biosignalverarbeitung'08*, 32–35.

[Koh08b] Kohl, F., Wübbeler, G., Kolossa, D., Orglmeister, R., Elster, C. and Bär, M. 2008. Performance of ICA for MEG data generated from subspaces with dependent sources. *Proc. ECIFMBE'08*, 22:1281–1285.

[Koh09] Kohl, F., Wübbeler, G., Kolossa, D., Elster, C., Bär, M., Orglmeister, R. 2009. Non-independent BSS: a model for evoked MEG signals with controllable dependencies. *Proc. ICA'09*, lecture notes in computer science, 5441:443–450.

[Koh10a] Kohl, F., Wübbeler, G., Kolossa, D., Bär, M., Orglmeister, R. and Elster, C. 2010. Shifted factor analysis for the separation of evoked dependent MEG signals. *Physics in medicine and biology*, 55:4219–4230.

[Koh10b] Kohl, F., Wübbeler, G., Kolossa, D., Elster, C., Bär, M. and Orglmeister, R. 2010. Noise adjusted PCA for finding the subspace of evoked dependent signals from MEG data. *Proc. LVA'10*, lecture notes in computer science, 6365:442–449.

[Kol09] Kolda, T.G. and Bader, B.W. 2009. Tensor decompositions and applications. *SIAM review*, 51(3):455–500.

[Kol06] Kolossa, D., Sawada, H., Astudillo, R.F., Orglmeister, R. and Makino, S. 2006. Recognition of convolutive speech mixtures by missing feature techniques for ICA. *Proc. ACSSC'06*, 1397–1401.

[Kop08] Kopriva, I. and Sersic, D. 2008. Wavelet packets approach to blind separation of statistically dependent sources. *Neurocomputing* 71(6-7):1642–1655.

[Köh00] Köhler, B.U. 2000. *Realzeitfähige blinde Quellentrennung am Beispiel elektroenzephalographischer Signale*, PHD thesis, Berlin, Technische Universität Berlin.

[Köh05] Köhler, B.U. 2005. *Konzepte der statistischen Signalverarbeitung*, Berlin, Springer.

[Kra04] Kraskov, A., Stögbauer, H. and Grassberger P. 2004. Estimating mutual information. *Phys. Rev. E*, 69(6):–066138.

[Kru77] Kruskal, J.B. 1977. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear algebra and its applications*, 18(2):95–138.

[Lat05] De Lathauwer, L., De Moor, B. and Vandewalle, J. 2005. A prewhitening-induced bound on the identification error in independent component analysis. *IEEE Transactions on circuits and systems I: regular papers*, 52(3):546–554.

[Lee90] Lee, J.B., Woodyatt, A.S. and Berman, M. 1990. Enhancement of high spectral resolution remote sensing data by a noise-adjusted principal component transform. *IEEE Trans. Geosci. Remote sensing*, 28(3):295–304.

[Lee99] Lee, T.W., Girolami, M. and Sejnowski, T.J. 1999. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural computation*, 11(2):417–441.

[Lem05] Lemm, S., Blankertz, B., Curio, G. and Müller, K.R. 2005. Spatio-spectral filters for improving the classification of single trial EEG. *IEEE Transactions on biomedical engineering*, 52(9):1541–1548.

[Lem06] Lemm, S., Curio,G., Hluschchuk,Y. and Müller, K.R. 2006. Enhancing the signal-to-noise ratio of ICA-based extracted ERPs. *IEEE Trans. Biomed. Eng.* 53(4):601–607.

[Lin09] Lin, F. 2009. *Separation of dependent components using subband independent component analysis*, master thesis, Berlin, Technische Universität Berlin.

[Lin05] Link, A., Burghoff, M., Salajegheh, A., Poeppel, D., Trahms, L. and Elster, C. 2005. Comparing a template approach and complex bandpass filtering for single-trial analysis of auditory evoked M100. *Biomedizinische Technik*, 52(1):106–110.

[Lü92] Lü, Z.L., Williamson, S.J. and Kaufman, L. 1992. Human auditory primary and association cortex have differing lifetimes for activation traces. *Brain research*, 572:236–241.

[Mak96] Makeig, S., Bell, A.J., Jung, T.P., Sejnowski, T.J. 1996. Independent component analysis of electroencephalographic data. *Proc. NIPS'96*, 8:145–51.

[Mak97] Makeig, S., Jung, T.P., Bell, A.J., Ghahremani, D. and Sejnowski, T.J. 1997. Blind separation of auditory event-related brain responses into independent components. *Proceedings of the national academy of sciences of the united states of america*, 94:10979–10984.

[Mak99a] Makeig, S., Westerfield, M., Jung, T.P., Covington, J., Townsend, J., Sejnowski, T.J. and Courchesne, E. 1999. Independent components of the late positive response complex in a visual spatial attention task. *Neuroscience*, 19(7):2665–2680.

[Mak99b] Makeig, S., Westerfield, M., Townsend, J., Jung, T.P., Courchesne, E., Sejnowski, T.J. 1999. Functionally independent components of early event-related potentials in a visual spatial attention task. *Philosophical transactions of the royal society of london. Series B: Biological Sciences*, 354:1135–1144.

[Mak02] Makeig, S., Westerfield, M., Jung, T.P., Enghoff, S., Townsend, J., Courchesne, E., Sejnowski, T.J. 2002. Dynamic brain sources of visual evoked responses. *Science*, 295:690–694.

[Mak04] Makeig, S., Debener, S., Onton, J. and Delorme, A. 2004. Mining event-related brain dynamics. *Trends in cognitive sciences*, 8(5):204–210.

[Mäk05] Mäkinen, V., Tiitinen, H. and May, P. 2005. Auditory event-related responses are generated independently of ongoing brain activity. *NeuroImage*, 24(4):961–968.

[Mäk88] Mäkelä, J.P., Hari, R. and Leinonen, L. 1988. Magnetic responses of the human auditory cortex to noise/square wave transitions. *Electroencephalography and clinical neurophysiology*, 69(5):423–430.

[McG76] McGurk, H. and MacDonald, J. 1976. Hearing lips and seeing voices. *Nature*, 264:746–748.

[McK98] McKeown, M. J., Makeig, S., Brown, G.G., Jung, T.P., Kindermann, S.S., Bell, A.J. and Sejnowski, T.J. 1998. Analysis of fMRI data by blind separation into independent spatial components. *Human brain mapping*, 6:160–188.

[Mei02] Meinecke, F., Ziehe, A., Kawanabe, M. and Müller, K.R. 2002. A resampling approach to estimate the stability of one-dimensional or multidimensional independent components. *IEEE transactions on biomedical engineering*, 49(12):1514–1525.

[Mis00] Miskin, J. 2000. *Ensemble learning for independent component analysis*, PhD thesis, University of Cambridge.

[Miw04] Miwakeichi, F., Martínez-Montes, E., Valdés-Sosa, P.A., Nishiyama, N., Mizuharam, H. and Yamaguchi, Y. 2004. Decomposing EEG data into space time frequency components using parallel factor analysis. *Neuroimage*, 22(3):1035–1045.

[Mol94] Molgedey, L. and Schuster, H.G. 1994. Separation of a mixture of independent signals using time delayed correlations. *Physical review letters*, 72(23):3634–3637.

[Mor05] Morup, M., Hansen, L.K., Herrmann, C.S., Parnas, J., Arnfred, S.M., 2005. Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG. *Neuroimage*, 29(3):938–947.

[Mor07] Morup, M., Madsen, K.H., Hansen, L.K., 2007. Shifted independent component analysis. *Proc. ICA'07*, 89–96.

[Mor08] Morup, M., Hansen, L.K., Arnfred, S.M., Lim, L.H. and Madsen, K.H. 2008. Shift-invariant multilinear decomposition of neuroimaging data. *Neuroimage*, 42(4):1439–1450.

[Möc86] Möcks, J. 1986. The influence of latency jitter in principal component analysis of event-related potentials. *Psychophysiology*, 23(4):480–484.

[Möc88] Möcks, J. 1988. Topographic components model for event-related potentials and some biophysical considerations. *IEEE transactions on biomedical engineering*, 35(6):482–484.

[Mül99] Müller, K.R., Philips, P. and Ziehe, A. 1999. JADE TD: combining higher-order statistics and temporal information for blind source separation (with noise). *Proc. ICA'99*, 1:87–92.

[Mül08] Müller, K.R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., Blankertz, B. 2008. Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. *Neuroscience methods*, 167(1):82–90.

[Mün00] Münte, T.F., Urbach, T.P., Düzel, E. and Kutas, M. 2000. Event-related brain potentials in the study of human cognition and neuropsycholgy. *Handbook of neuropsychology*, 1:139–235.

[Nää79] Näätänen, R. and Michie, P.T. 1979. Early selective-attention effects on the evoked potential: a critical review and reinterpretation. *Biological psychology*, 8(2):81–136.

[Nää87] Näätänen, R. and Picton, T. 1988. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, 24(4):375–425.

[Nää94] Näätänen, R., Ilmoniemi, R.J. and Alho, K. 1994. Magnetoencephalography in studies of human cognitive brain function. *Trends in neuroscience*, 17(9): 389–395.

[Nie05] Niedermeyer, E. and Lopes Da Silva, F.H. 2005. *Electroencephalography: basic principles, clinical applications, and related fields.* Philadelphia, Lippincott Williams & Wilkins.

[Nie09] Nierhaus, T., Schön, T., Becker, R., Ritter, P. and Villringer, A. 2009. Background and evoked activity and their interaction in the human brain. *Magnetic resonance imaging*, 27(8):1140–1150.

[Nik93] Nikias, C.L. and Petropulu, A.P. 1993. *Higher-order spectra analysis:A nonlinear signal processing framework.*

[Nol04] Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S. and Hallett, M. 2004. Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical Neurophysiology*, 115(10):2292–2307.

[Nun81] Nunez, P.L. 1981. *Electric Fields of the Brain: The Neurophysics of EEG*, New York, Oxford University.

[Nuw98] Nuwer, M.R. 1998. Fundamentals of evoked potentials and common clinical applications today. *Electroencephalography and clinical neurophysiology*, 106(2):142-148.

[Oga90] Ogawa, S., Lee, T.M., Nayak, A.S. and Glynn, P. 1990. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic resonance in medicine*, 14(1):68–78.

[Oka82] Okada, Y.C., Kaufman, L., Brenner, D. and Williamson, S.J. 1982. Modulation transfer functions of the human visual system revealed by magnetic field measurements. *Vision research*, 22(2):319–333.

[Ont06] Onton, J., Westerfield, M., Townsend, J. and Makeig, S. 2006. Imaging human EEG dynamics using independent component analysis. *Neuroscience & biobehavioral reviews*, 30(6):808–822.

[Paa00] Paatero, P. 2000. Construction and analysis of degenerate parafac models. *Journal of chemometrics*, 14(3):285–299.

[Pan88] Pantev, C., Hoke, M., Lehnertz, K., Lütkenhöner, B., Anogianakis, G. and Wittkowski, W. 1988. Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroencephalography and clinical neurophysiology*, 69(2):160–170.

[Pap07] Papoulis, A., Pillai, S.U. and Unnikrishna, S. 2007. *Probability, random variables, and stochastic processes.* New York, McGraw-Hill.

[Par03] Parra, L., Alvino, C., Tang, A., Pearlmutter, B., Yeung, N., Osman, A. and Sajda, P.2003. Single-trial detection in EEG and MEG: keeping it linear. *Neurocomputing*, 52-54:177–183.

[Pfu99] Pfurtscheller, G. and Lopes da Silva, F.H. 1999. Event related EEG/MEG synchronisation and desynchronisation : basic principles. *Clinical Neurophysiology*, 110(11):1842–1857.

[Pic95] Picton, T.W. and Lins, O.G. and Scherg, M. 1995. The recording and analysis of event-related potentials. *Handbook of neuropsychology*, 10:3–73.

[Poc05] Póczos, B. and Lörincz,A. 2005. "Independent subspace analysis using k-nearest neighborhood distances". in *Proc. ICANN'05*, pp. 163-168, 2005.

[Poc06] Póczos, B. and Lörincz, A. 2006. Non-combinatorial estimation of independent autoregressive sources. *Neurocomputational Letters* 69(16-18):2416–2419.

[Raj08] Rajih, M., Comon, P. and Harshman, R.A. 2008. Enhanced line search: a novel method to accelerate PARAFAC. *SIAM journal on matrix analysis and applications*, 30(3):1128–1147.

[Rao71] Rao, C.R. and Mitra, S. 1971. *Generalized inverse of matrices and its applications*, New York, Wiley.

[Roy90] Roy, C.S. and Sherrington, C.S. 1890. On the regulation of the blood-supply of the brain. *Physiology*, 11(1-2):85–108.

[Sal04] Salajegheh, A., Link, A., Elster, C., Burghoff, M., Sander, T., Trahms, L. and Poeppel, D. 2004. Systematic latency variation of the auditory evoked M100: from average to single-trial. *Neuroimage* 23(1):288–295.

[San10] Sander, T.H., Knösche, T.R., Schlögl, A., Kohl, F., Wolters, C.H., Haueisen, J. and Trahms, L. 2010. Recent advances in modeling and analysis of bioelectric and biomagnetic sources. *Biomedizinische Technik*, 55:65–76.

[Sar87] Sarvas, J. 1987. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Physics in medicine and biology*, 32(1): 11–22.

[Say74] Sayers, B.M.A., Beagley, H.A. and Henshall, W.R. 1974. The mechanism of auditory evoked EEG responses. *Nature*, 247:481–483.

[Sch91] Scharf, L.L. and Demeure, C 1991 *Statistical signal processing: detection, estimation, and time series analysis*, Massachusetts, Addison-Wesley.

[Sha48] Shannon, C.E. 1948. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423 and 623–656.

[Sha00] Shawkat, F.S. and Kriss, A. 2000. A study of the effects of contrast change on pattern VEPs, and the transition between onset, reversal and offset modes of stimulation. *Documenta ophthalmologica*, 101(1):73–89.

[Ski53] Skitovitch, V.P. 1953. On a property of the normal distribution. *Doklady akademii nauk SSSR*, 89:217–219.

[Slo99] Slotnick, S.D., Klein, S.A., Carney, T., Sutter, E. and Dastmalchi, S. 1999. Using multi-stimulus VEP source localization to obtain a retinotopic map of human primary visual cortex. *Clinical neurophysiology*, 110(10):1793-1800.

[Sti98] Stinstra, J.G. and Peters, M.J. 1998. The volume conductor may act as a temporal filter on the ECG and EEG. *Medical and Biological Engineering and Computing*, 36(6):711–716.

[Sto02] Stone, J.V., Porrill, J., Porter, N.R. and Wilkinson, I.D. 2002. Spatiotemporal independent component analysis of event-related fmri data using skewed probability density functions. *Neuroimage*, 15:407–421.

[Str80] Strang, G. 1980. *Linear Agebra and Its Applications.* Philadelphia, Academic Press.

[Sun04] Sun, L., Shen, M., Xu, W. and Chan, F. 2004. Single-trial estimation of multi-channel VEP signals using independent component analysis. *Proc. ISNN'04*, lecture notes in computer science, 3173:732–737.

[Sza06] Szabó, Z.,Pószos B., and Lörincz A., 2006. Separation theorem for k-independent subspace analysis with sufficient conditions. *Technical report*, Eötvös Lorand University, Budapest.

[Tan04] Tanaka, T. and Cichocki, A. 2004. Subband decomposition independent component analysis and new performance criteria. *Proc. ICASSP'04*, 5:541-544.

[Tan02] Tang, A.C., Pearlmutter, B.A., Malaszenko, N.A., Phung, D.B. and Reeb, B.C. 2002. Independent components of magnetoencephalography: localization. *Neural Computation*, 14(18):1827–1858.

[Tan05] Tang, A.C., Liu, J.Y. and Sutherland, M.T. 2005. Recovery of correlated neuronal sources from EEG: the good and bad ways of using SOBI. *Neuroimage*, 28:507–519.

[Ter75] Ter-Pogossian, M.M., Phelps, M.E., Hoffman, E.J. and Mullani, N.A. 1975. A positron-emission transaxial tomograph for nuclear imaging (PETT). *Radiology*,114(1):89–98.

[Tha93] Thakor, N.V., Xin-Rong, G.,Yi-Chun, S. and Hanley, D.F. 1993. Multiresolution wavelet analysis of evoked potentials. *IEEE transactions on biomedical engineering*, 40(11):1085–1094.

[The05] Theis, F.J., Gruber, P., Keck,I.R., Meyer-Baese A. and Lang. E.W. 2005. Spatiotemporal blind source separation using double-sided approximate joint diagonalization. *Proc. EUSIPCO'05*, Antalya, Turkey.

[The06] Theis, F.J. 2006. Towards a general independent subspace analysis. *Proc. NIPS'06*, 19:1361–1368.

[Tru02] Truccolo, W.A., Ding, M., Knuth, K.H., Nakamura, R. and Bressler, S.L. 2002. Trial-to-trial variability of cortical evoked responses: implications for the analysis of functional connectivity. *Clinical neurophysiology*, 113(2):206–226.

[Vig97] Vigário, R.N., Jousmáki, V., Hämäläinen, M., Hari, R. and Oja, E. 1998. Independent component analysis for identification of artifacts in magnetoencephalographic recordings. *Proc. NIPS'98*, 10:229–235.

[Vig00] Vigário, R.N. 2000. Dipole modeling in FastIca decomposition of evoked responses. *Proc. ICA'00*, 423–428.

[Vig09] Vigário, R.N. and Oja, E. 2008. BSS and ICA in neuroinformatics: from current practices to open challenges. *IEEE Reviews in biomedical engineering*, 1:50–61.

[Wal05] Walsh, P.,Kane, N. and Butler, S. 2005. The clinical role of evoked potentials. *Journal of neurology, neurosurgery & psychiatry*, 76(2):16–22.

[Wan00] Wang, K., Begleiter, H. and Porjesz, B. 2000. Trilinear modeling of event-related potentials. *Brain topography*, 12(4):263–271.

[Wil06] Wilkinson, A.R. and Jiang, Z.D. 2006. Brainstem auditory evoked response in neonatal neurology. *Seminars in Fetal and Neonatal Medicine,* 11(6):444–451.

[Woo67] Woody, C.D. 1967. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Medical and biological engineering and computing*, 5(6):539–554.

[Wüb07] Wübbeler, G., Link, A., Burghoff, M., Trahms, L. and Elster, C. 2007. Latency analysis of single auditory evoked M 100 responses by spatio-temporal filtering. *Physics in medicine and biology*, 52:4383–4392.

[Xu09] Xu, L., Stoica, P., Li, J., Bressler, S.L., Shao, X. and Ding, M. 2009. ASEO: a method for the simultaneous estimation of single-trial event-related potentials and ongoing brain activities. *IEEE transactions on biomedical engineering*, 56(1):111–121.

[Yer02] Yeredor, A. 2002. Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation. *IEEE Transactions on signal processing*, 50(7):1545–1553.

[Yli08] Ylipaavalniemi, J. and Vigário, R. 2008. Analyzing consistency of independent components: An fMRI illustration. *Neuroimage*, 39(1):169–180.

[Zha06a] Zhang, K. and Chan, L.W. 2006. Enhancement of source independence for blind source separation. *Proc. ICA'06*, lecture notes in computer science, 3889:731–738.

[Zha06b] Zhang, K. and Chan, L.W. 2006. An adaptive method for subband decomposition ICA. *Neural computation*, 18(1):191-223.

[Zie98] Ziehe, A. and Müller, K.R. 1998. TDSEP-an efficient algorithm for blind separation using time structure. *Proc. ICANN'98*, 675–680.

[Zim77] Zimmerman, J.E. 1977. SQUID instruments and shielding for low-level magnetic measurements. *Journal of applied physics*, 48:702–710.

[Zav09] Zavala-Fernandez, H. 2009. *Evaluation and comparison of the independent components of simultaneously measured MEG and EEG Data*, PHD thesis, Berlin, Technische Universität Berlin.

[Zsc02] Zschocke, S. 2002. *Klinische Elektroenzephalographie,* Berlin, Springer.

Appendix

Audio-visual paradigma code using the software *Presentation*®

```
# header
scenario = "AVAV2";
scenario_type = trials;
no_logfile = false;
default_background_color = 255, 255, 255;
write_codes = true;
pulse_width = 30;
active_buttons = 1;

# port_code:
# 10  —— A
# 20  —— V
# 40  —— AV_0
# 40–85  —— AV_0_20

# trial definitions
begin;#
# white picture
picture { bitmap { filename = "weiss2.png"; height = 300; width = 450;}; x = 0; y = 0;} pic_weiss;
# resting picture
picture { bitmap { filename = "rest.jpg"; height = 300; width = 450;} ; x = 0; y = 0;} pic_rest;
# auditory tone of 1 kHz / duration 300 ms
sound { wavefile { filename = "1000Hzsound.wav"; }; }  tone_1kHz;
# visual picture checherwedge 1
picture { bitmap { filename = "stim_wedge_1.png"; height = 300; width = 450;}; x = 0; y = 0;} pic_check1;
# visual picture checherwedge 2
picture { bitmap { filename = "stim_wedge_2.png"; height = 300; width = 450;}; x = 0; y = 0;} pic_check2;

trial {  nothing {}; port_code=10; picture pic_weiss; stimulus_event{sound tone_1kHz; } A_event; } A_trial;
trial {  stimulus_event{picture pic_check1; } V1_event; stimulus_event{picture pic_check2; } V2_event; } V_trial;
trial {  stimulus_event{picture pic_check1; } AV_V1_event;  stimulus_event{picture pic_check2; } AV_V2_event;
                                                            stimulus_event{sound tone_1kHz; }    AV_A_event; } AV_trial;
# buffer, rest, attention trial
trial {  trial_duration=1300; picture pic_weiss; code="buffer"; } buffer_trial;
trial {  picture pic_rest; duration=response; code="rest"; } resting_trial;

# PCL Code starts here

begin_pcl;
int ran_val;
int ran_jit;
int main_exp;
int part;
int fixdelay;
int genshift;
```

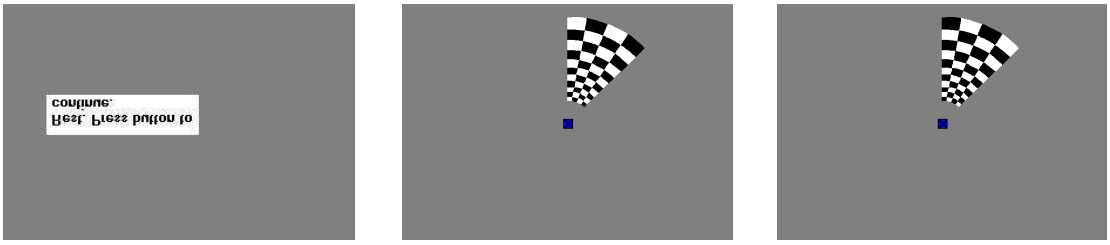


Figure 6.1: Original paradigma pictures: resting, checkerwedge 1 and 2. Mirrors reflect the picture from the beamer to the screen and the subjects sees mirrored versions.

```
string proband;
string strme;
string strmea;
int strmeb;
int tog;
int i;
int ii;
int j ;
fixdelay=50;
proband="Florian Kohl";
main_exp=0;
genshift=0;
part=1;

# Start of paradigma
if (main_exp<100) then

strme=proband;
strme.append("mainexp");
strme.append("-");
strme.append(string(main_exp));
strme.append("-");
strme.append(string(part));
strme.append(".log");
logfile.set_filename(strme);

j = 1;
loop until j > 5
begin

# auditory trials
i = 1;
loop until i > 8
begin
   ran_val = random(1500, 2000);
   A_trial.set_duration(ran_val);
   A_event.set_time(fixdelay+genshift);
   A_event.set_event_code( "tone_1kHz_A_trial" );
   A_trial.present();
   i=i+1;
end;

# visual trials
i = 1;
tog=1;
loop until i > 11
begin
   ran_val = random(1500, 2000);
   V_trial.set_duration(ran_val);
   if tog==1 then
   V1_event.set_stimulus(pic_check1);
   V2_event.set_stimulus(pic_check2);
   tog=0;
   else
   V1_event.set_stimulus(pic_check2);
   V2_event.set_stimulus(pic_check1);
   tog=1;
   end;
   V1_event.set_port_code(20);
   V2_event.set_deltat(fixdelay-5);
   V1_event.set_event_code( "V1_V_trial" );
   V2_event.set_event_code( "V2_V_trial" );
   V_trial.present();
   i=i+1;
end;

ii = 1;
loop until ii > 2
begin

# audio-visual with 17ms time shifts
i = 1;
loop until i > 3
begin
   ran_val = random(1500, 2000);
   AV_trial.set_duration(ran_val);
   if tog==1 then
   AV_V1_event.set_stimulus(pic_check1);
   AV_V2_event.set_stimulus(pic_check2);
   tog=0;
   else
   AV_V1_event.set_stimulus(pic_check2);
   AV_V2_event.set_stimulus(pic_check1);
   tog=1;
   end;
   ran_jit = random(1, 3);
   if (ran_jit==1) then      AV_V1_event.set_port_code(40);
        AV_V1_event.set_event_code( "AV_jit0" );
```

```
        V_V2_event.set_deltat(fixdelay -5);
        AV_A_event.set_deltat(genshift); end;
   if (ran_jit==2) then       AV_V1_event.set_port_code(60);
        AV_V1_event.set_event_code( "AV_jit16_1" );
        AV_V2_event.set_deltat(fixdelay -5-17);
        AV_A_event.set_deltat(genshift+17); end;
   if (ran_jit==3) then       AV_V1_event.set_port_code(65);
        AV_V1_event.set_event_code( "AV_jit16_2" );
        AV_V2_event.set_deltat(fixdelay -5+17);
        AV_A_event.set_deltat(genshift -17); end;
   AV_A_event.set_event_code( string(ran_jit));
   AV_V2_event.set_event_code( "V_AV_trial" );
   AV_trial.present();
   i=i+1;
end;

# audio-visual with 0ms time shifts
i = 1;
loop until i > 3
begin
   ran_val = random(1500, 2000);
   AV_trial.set_duration(ran_val);
   if tog==1 then
   AV_V1_event.set_stimulus(pic_check1);
   AV_V2_event.set_stimulus(pic_check2);
   tog=0;
   else
   AV_V1_event.set_stimulus(pic_check2);
   AV_V2_event.set_stimulus(pic_check1);
   tog=1;
   end;
   AV_V1_event.set_port_code(40);
        AV_V1_event.set_event_code( "AV_jit0" );
        AV_V2_event.set_deltat(fixdelay -5);
        AV_A_event.set_deltat(genshift);
        AV_A_event.set_event_code( string(ran_jit));
   AV_V2_event.set_event_code( "V_AV_trial" );
   AV_trial.present();
   i=i+1;
end;


# audio-visual with 8ms time shifts
i = 1;
loop until i > 3
begin
   ran_val = random(1500, 2000);
   AV_trial.set_duration(ran_val);
   if tog==1 then
   AV_V1_event.set_stimulus(pic_check1);
   AV_V2_event.set_stimulus(pic_check2);
   tog=0;
   else
   AV_V1_event.set_stimulus(pic_check2);
   AV_V2_event.set_stimulus(pic_check1);
   tog=1;
   end;
   ran_jit = random(1, 3);
   if (ran_jit==1) then       AV_V1_event.set_port_code(40);
        AV_V1_event.set_event_code( "AV_jit0" );
        AV_V2_event.set_deltat(fixdelay -5);
        AV_A_event.set_deltat(genshift); end;
   if (ran_jit==2) then       AV_V1_event.set_port_code(80);
        AV_V1_event.set_event_code( "AV_jit8_1" );
        AV_V2_event.set_deltat(fixdelay -5);
        AV_A_event.set_deltat(genshift -8); end;
   if (ran_jit==3) then       AV_V1_event.set_port_code(85);
        AV_V1_event.set_event_code( "AV_jit8_2" );
        AV_V2_event.set_deltat(fixdelay -5);
        AV_A_event.set_deltat(genshift+8); end;
   AV_A_event.set_event_code( string(ran_jit));
   AV_V2_event.set_event_code( "V_AV_trial" );
   AV_trial.present();
   i=i+1;
end;
ii=ii+1;
end; # jitter loop

resting_trial.present();
buffer_trial.present();
j=j+1;
end;
```