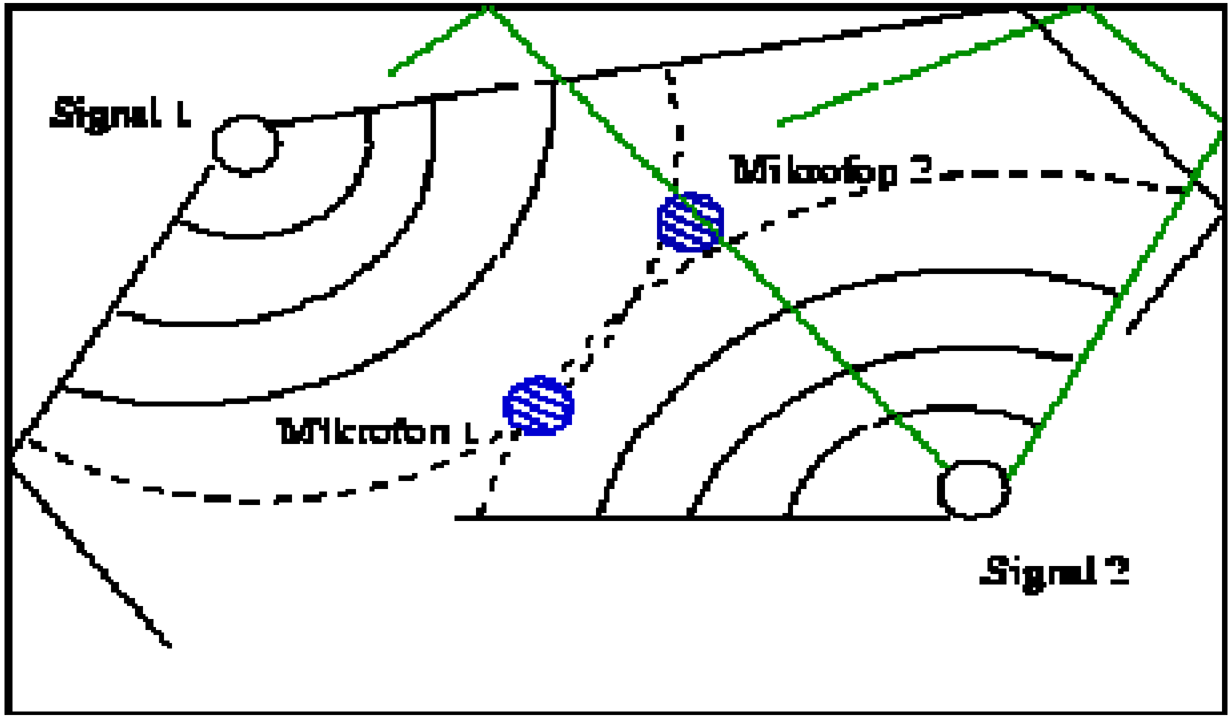
Independent Component Analysis (ICA)

Klaus-Robert Müller

Blind Source Separation



applications:

cocktailparty problem, biomedical measurements (EEG, MEG), etc.

question:

decomposition & analysis of **superimposed** signals, robust denoising.

Acoustic Demo: "Cocktail party"

- 3 mixed signals (music, speech, street noise)

  $$\mathbf{x}(t) = A\mathbf{s}(t)$$

- problem: **demixing**!

“Blind” Source Separation I

microphones $\mathbf{x}(t)$ measure **unknown** mixtures of **unknown** (sound) sources

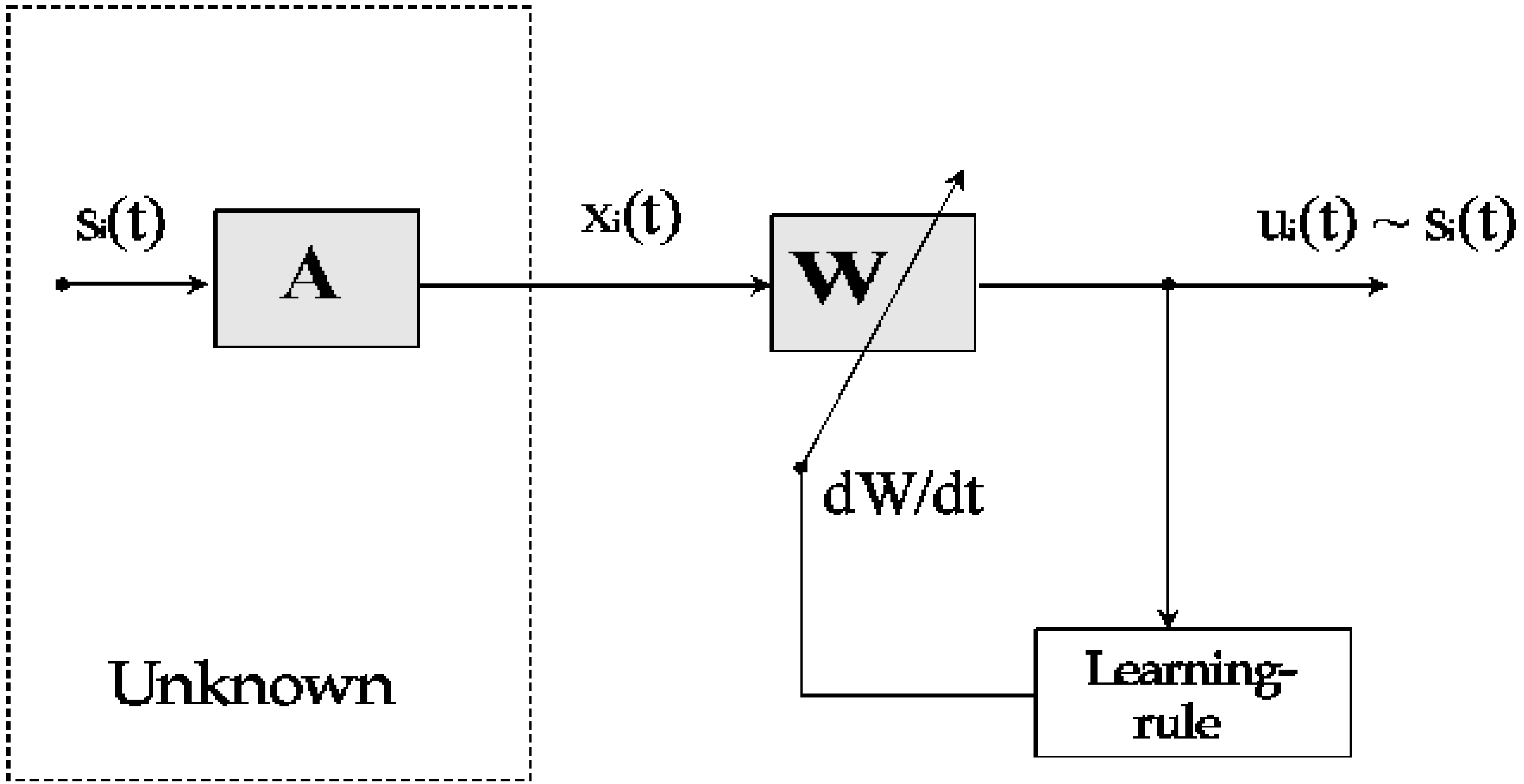
$$\mathbf{x}(t) = \mathbf{A}\,\mathbf{s}(t)$$

assumption: statistical independence of the source signals (**ICA**)

Ansatz: invert mixing process **A** by **learning** of **W** and enforce statistical independence of unmixed signals $\mathbf{u}$(t)!

$$\mathbf{u}(t) = \mathbf{W}\,\mathbf{x}(t)$$

“Blind” Source Separation II



“Blind” Source Separation III

assumption: statistical independence of sources

$$p(\boldsymbol{u}) \;=\; \prod_{i=1}^{n} p_i(u_i)$$

•higher cross-moments should vanish

•minimize distance between distributions

$$D(\mathbf{W}) = \int p(\boldsymbol{u}) \log \left(\frac{p(\boldsymbol{u})}{\prod_{i=1}^{n} p_i(u_i)} \right) d\boldsymbol{u}$$

problem: how to obtain distributions in practice?

“Blind” Source Separation IV

Gram Chalier expansion:

$$p_i(u_i) \sim \frac{1}{\mathcal{N}} e^{-(u_i)^2/2} \left(1 + \frac{m_i^{(3)}}{3!} H_3(u_i) + \frac{[m_i^{(4)} - 3]}{4!} H_4(u_i) + \dots \right)$$

Edgeworth expansion:

$$\begin{aligned} p_i(u_i) \quad \sim & \quad \frac{1}{\mathcal{N}} e^{-(u_i)^2/2} (1 + \frac{m_i^{(3)}}{3} H_3(u_i) + \frac{m_i^{(4)}}{4} H_4(u_i) + \\ + & \quad \frac{10}{6} (m_i^{(3)})^2 H_6(u_i) + \frac{1}{5} m_i^{(5)} H_5(u_i) + \frac{35}{8} m_i^{(3)} m_i^{(4)} H_7(u_i) \\ + & \quad \dots) \end{aligned}$$

where $m_i^{(k)}$ is kth order moment of u_i and $H_k(u_i)$ are Chebyshev-Hermite polynomials (order k).

“Blind” Source Separation V

after tedious but straight forward calculation, we get

$$
\begin{aligned}
D(\mathbf{W}) \quad \sim \quad & -\int p(\boldsymbol{x}) \log(p(\boldsymbol{x})) - \log \| \det(\mathbf{W}) \| + \frac{n}{2} \log(2\pi e) + \dots \\
& -\sum_{i=1}^{n} [\frac{(m_i^{(3)})^2}{2 \cdot 3!} + \frac{[m_i^{(4)} - 3]^2}{2 \cdot 4!} - \frac{5}{8} (m_i^{(3)})^2 [m_i^{(4)} - 3] + \dots \\
& - \frac{1}{16} [m_i^{(4)} - 3]^3]
\end{aligned}
$$

gradient descent in D(W) with respect to W yields

$$\frac{d\mathbf{W}}{dt} = \eta(t) \{ \mathbf{I} - \mathbf{f}(\boldsymbol{u}) \boldsymbol{u}^T \} \mathbf{W} \qquad \text{(e.g. Amari et al. 96)}$$

$$f(u) = 3/4u^{11} + 25/4u^9 - 47/4u^5 + 29/4u^3.$$

“Blind” Source Separation with Temporal Information

model: $\mathbf{x}(t) = \mathbf{A}\,\mathbf{s}(t)$, $\mathbf{u}(t) = \mathbf{W}\,\mathbf{x}(t)$

define covariance matrices over time:

$$\mathbf{V} = \langle \boldsymbol{x}_t \boldsymbol{x}_t^T \rangle \qquad \mathbf{V}_\tau = \langle \boldsymbol{x}_t \boldsymbol{x}_{t-\tau}^T \rangle \qquad \forall\, i \neq j,$$

assumption: s has significant autocorrelation

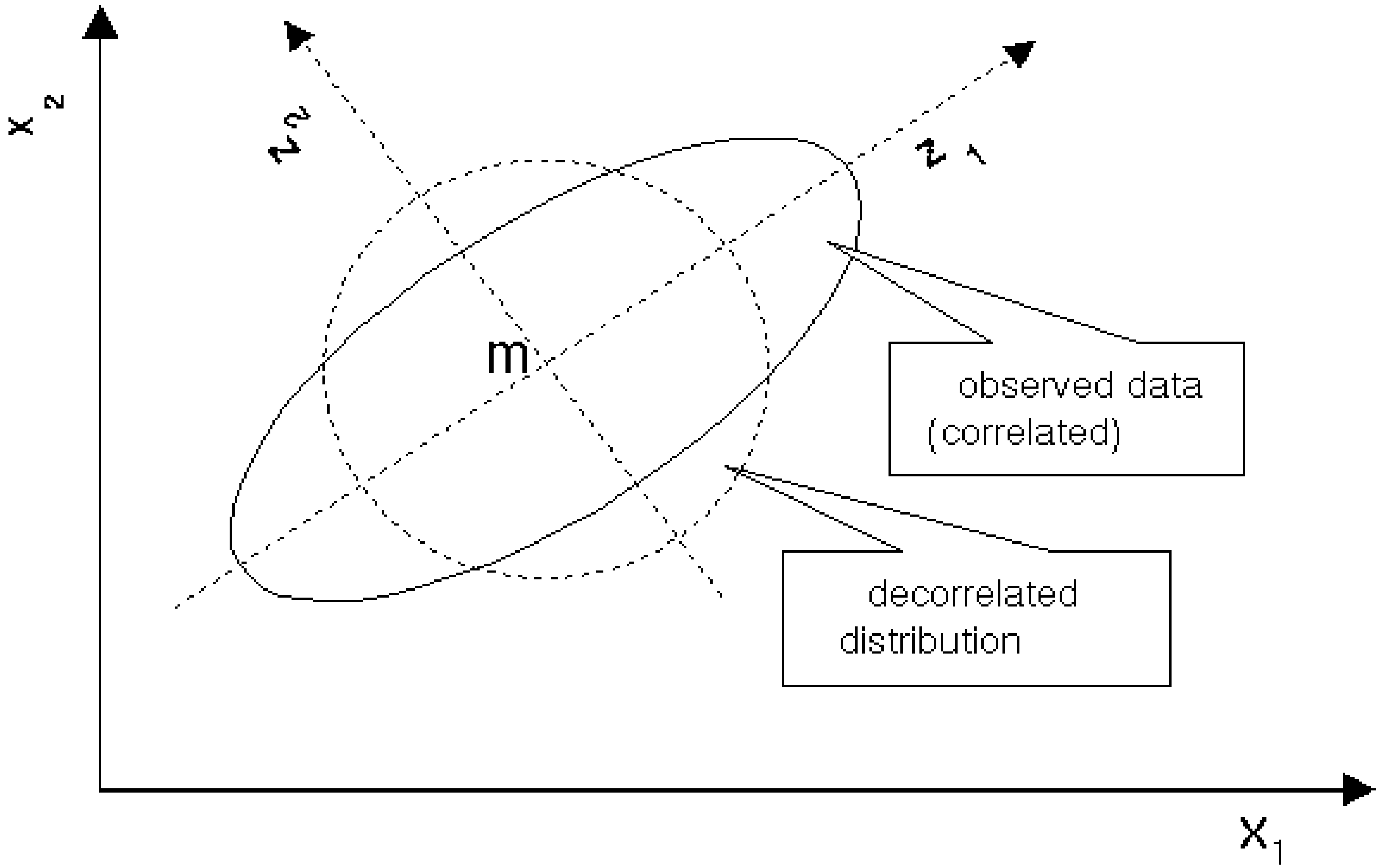
algorithm: **TDSEP** minimizes error

$$L\{\mathbf{W}\} = \sum_{i \neq j} \langle u_i(t) u_j(t) \rangle^2 + \sum_{\{\tau\}} \langle u_i(t) u_j(t-\tau) \rangle^2$$

solution: linear algebra vs. gradient descent

simultaneous diagonalisation of $\{\mathbf{V},\ \mathbf{V}_\tau, \ldots\}$

Whitening and Jacobi Rotations I



Whitening and Jacobi Rotations II

whitening transformation **K** is e.g. determined as inverse square root of the covariance matrix

$$\mathbf{K} = \langle \boldsymbol{x}\boldsymbol{x}^T \rangle^{-\frac{1}{2}} = (\boldsymbol{v}\,\Lambda\,\boldsymbol{v}^T)^{-\frac{1}{2}} = \boldsymbol{v}\,\Lambda^{-\frac{1}{2}}\,\boldsymbol{v}^T.$$

then approximative simultaneous diagonalisa-tion of transformed time-delayed covariance matrix

$$\mathbf{V}_{\tau(z)} \quad = \quad \langle z_t z_{t-\tau}^T \rangle = \boldsymbol{Q}^T\,\mathbf{V}_{\tau(s)}\,\boldsymbol{Q} = \boldsymbol{Q}^T \Lambda_\tau\,\boldsymbol{Q}.$$

solution: $\mathbf{A} = \mathbf{K}^{-1}\boldsymbol{Q}$

How to enforce statistical independence?

model: **x**(t) **= A s**(t) **u**(t) **= W x**(t)

higher order statistics (expansions)

$$\frac{d\mathbf{W}}{dt} = \eta(t)\{\mathbf{I} - \mathbf{f}(\boldsymbol{u})\boldsymbol{u}^T\}\mathbf{W} \qquad \text{(e.g. Amari et al. 96)}$$

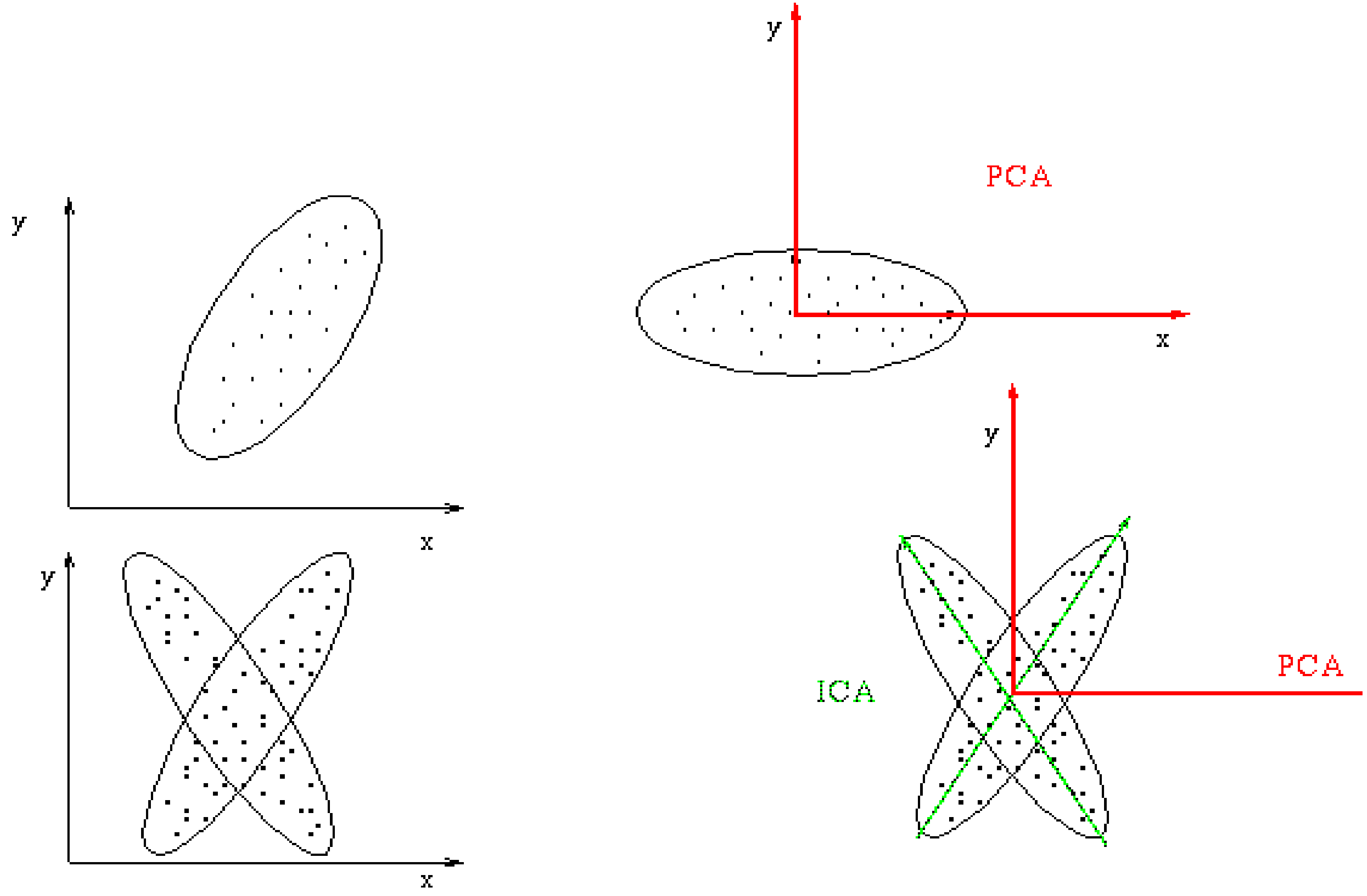
$$f(u) = 3/4u^{11} + 25/4u^9 - 47/4u^5 + 29/4u^3.$$

second order statistics & **temporal information**

$$L\{\mathbf{W}\} = \sum_{i \neq j} \langle u_i(t) u_j(t) \rangle^2 + \sum_{\{\tau\}} \langle u_i(t) u_j(t-\tau) \rangle^2$$

(simultaneous diagonalisation of matrices, TDSEP)

PCA vs ICA



Acoustic Demo II

- 3 mixed signals (music, speech, street noise)

$$\mathbf{x}(t) = A\mathbf{s}(t)$$

- problem: music signal has **very small** amplitude, i.e. hidden signal
- question: which music instrument?

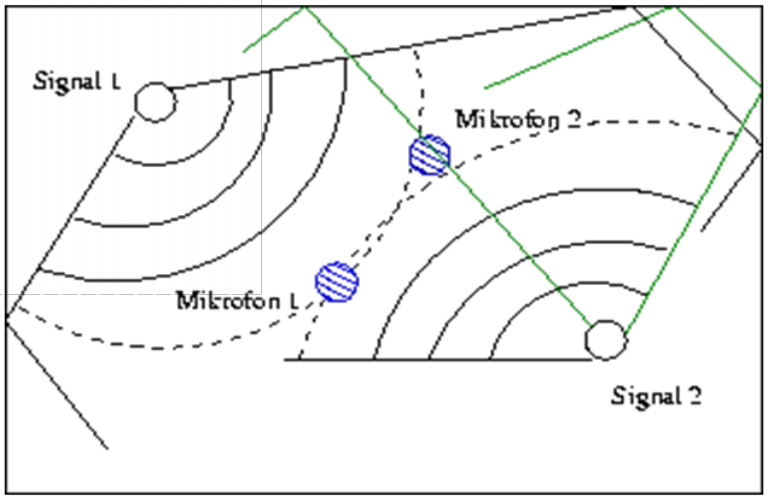
- *mixed* *unmixed signal*

Cf. cerebral cocktail party problem

Acoustic Demo III

- 2 mixed signals (real room = convolution)

$$x_i(t) = \sum_{j=1}^{N}\sum_{\tau=0}^{M} h_{ij}(\tau)s_j(t-\tau),$$
$$= \sum_{j=1}^{N} h_{ij} * s_j(t),$$

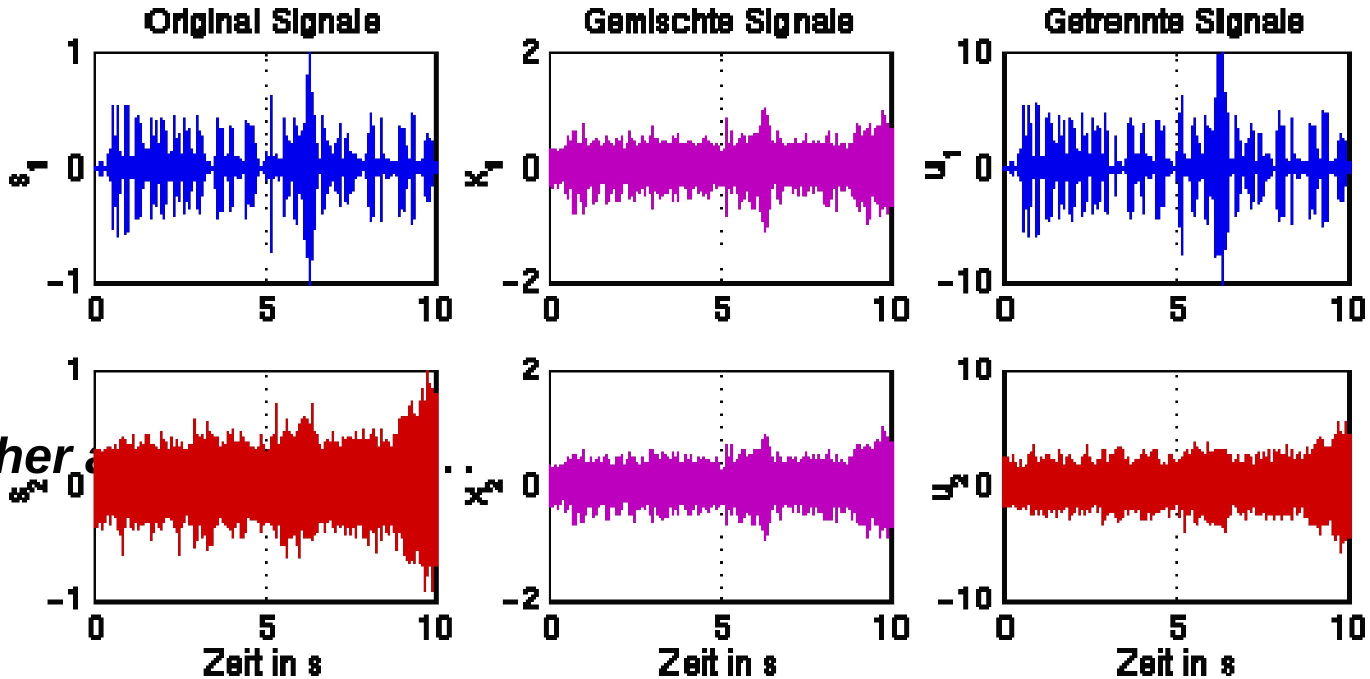


- *mixed*

- *unmixed signal*

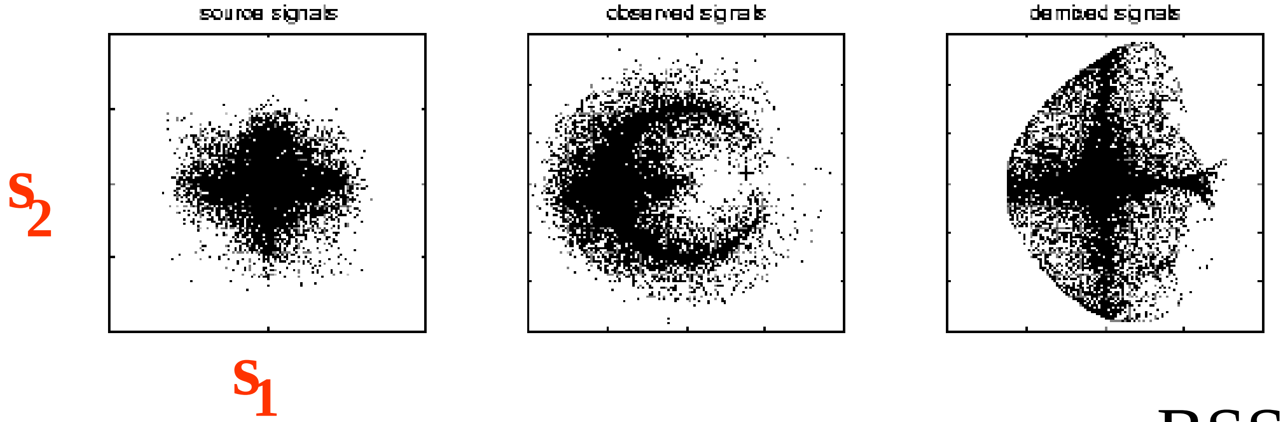
Acoustic Demo IV

- adaptive source separation: 2 mixtures
- *mixed* *unmixed signal*

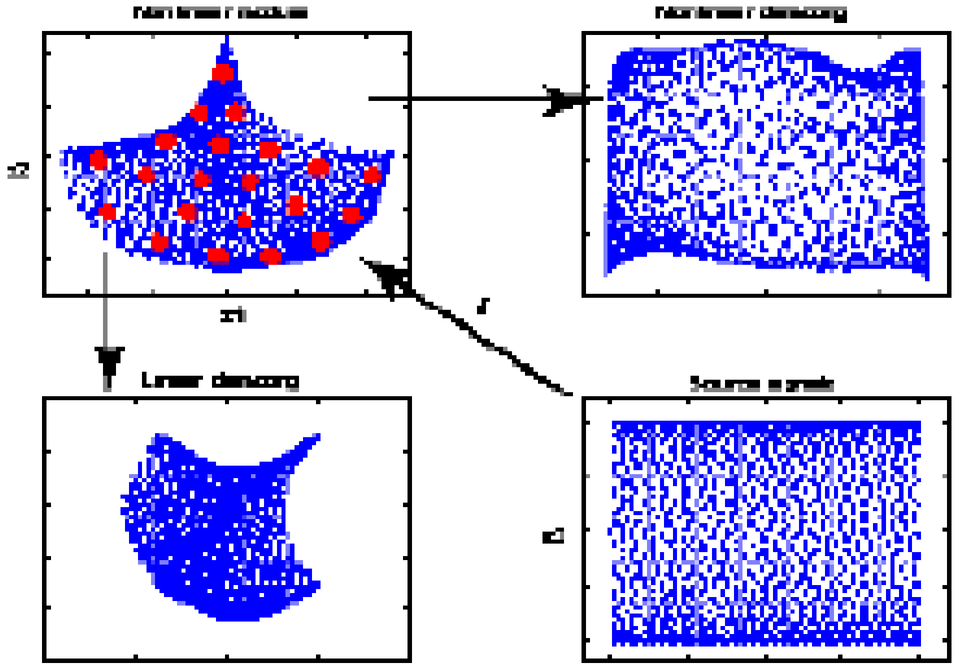


further a

Nonlinear source separation



$$\mathbf{x}[t] = f(\mathbf{s}[t])$$



BSS of **nonlinearly** distorted mixtures with kernel based learning methods
[Harmeling et al. 2001, 2002, Ziehe et al. 2001]

$$\mathbf{x}[t] = f(A\mathbf{s}[t])$$

An ICA Analysis of Non-invasively Recorded DC-fields in Humans

Klaus-Robert Müller, Andreas Ziehe, Gerd Wübbeler,

Bruno-Marcel Mackert, Lutz Trahms, Gabriel Curio

Fraunhofer FIRST Berlin & Universität Potsdam, klaus@first.fhg.de

AG Neurophysics, Charite Berlin

and PTB Berlin

Cortical Signals

- brain works distributed and parallel
- idea: discriminate “speakers in brain”
- signal processing problem analog to

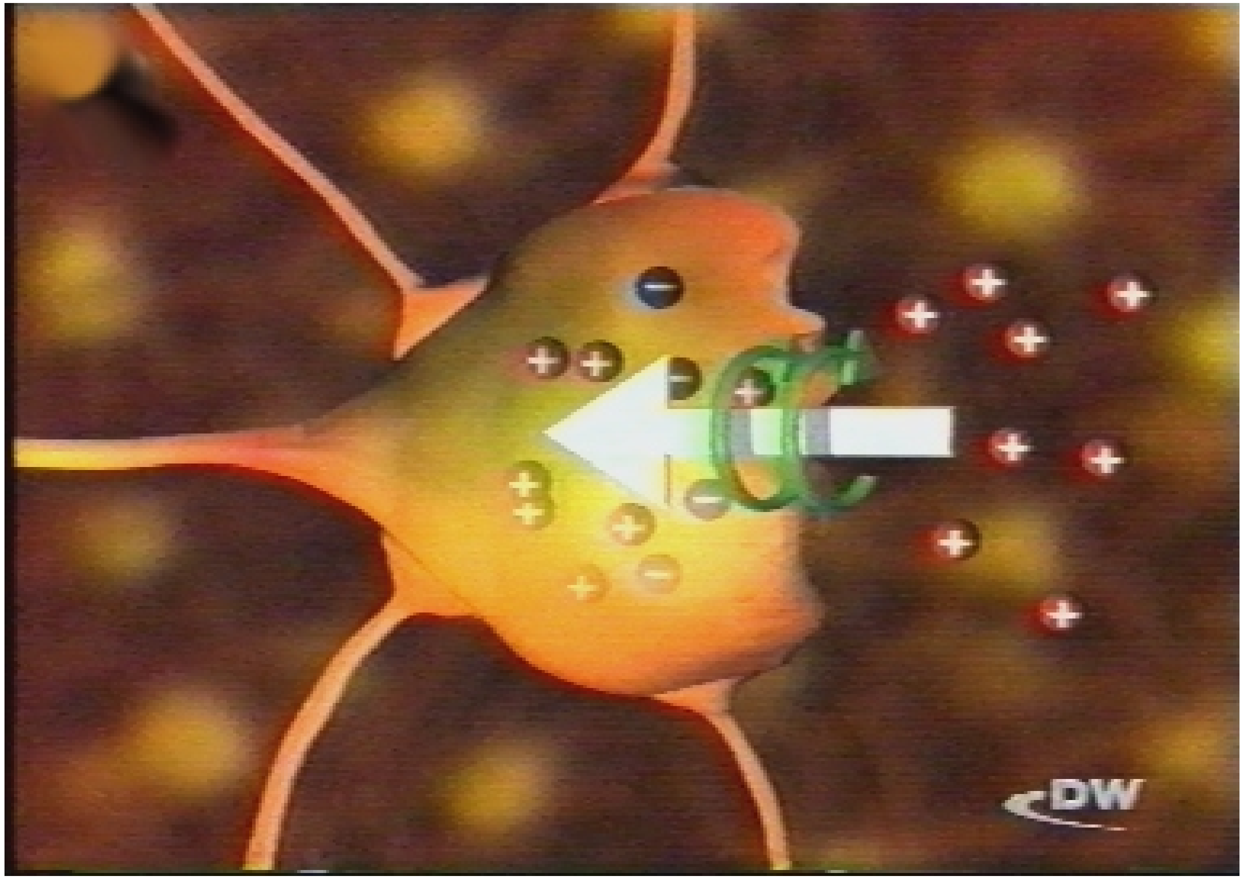
  Cocktailparty problem

Cortical Signals II

- ***GOAL***: identification and extraction of small brain signals despite of noise (external or physiological “noise”, i.e. background activity)
- denoised signals as basis for neurophysiological modeling
- challenge for signal processing, time series prediction and machine learning
- reliability of the analysis
- relevant signals are often extremely weak compared to the noise, i.e. a factor of 10000!

Setup: shielded MEG chamber

Why do we measure magnetic fields?



Magnetic fields show brain activity:
single neurons depolarize → synchronous active neuron populations alow a non-invasive monitoring of macroscopic currents.

Setup II

MEG positioning near the auditory cortex

Analysis of DC MEG

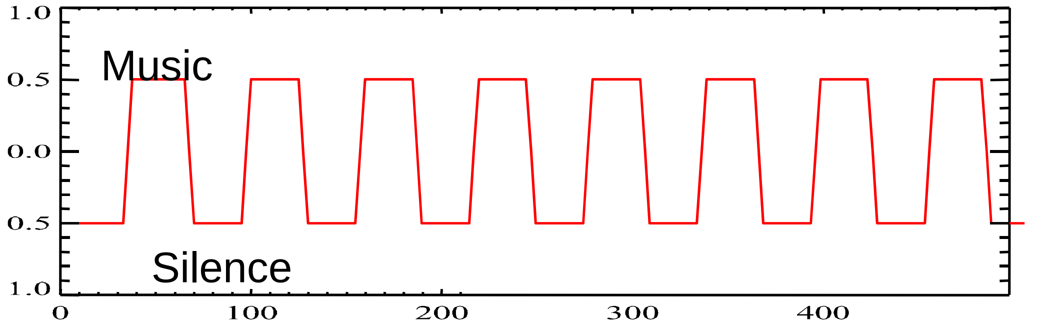
Paradigm: acoustic stimulation by presentation of alternating periods of music and silence, each for 30 s;

Non-invasive measurements of magnetic fields over the left auditory cortex for 30 min with 49 channel SQUID gradiometer

mechanical horizontal modulation of the body position with a frequency of 0.4 Hz,

- transposes DC magnetic field into higher frequency to improve the signal-to-noise ratio

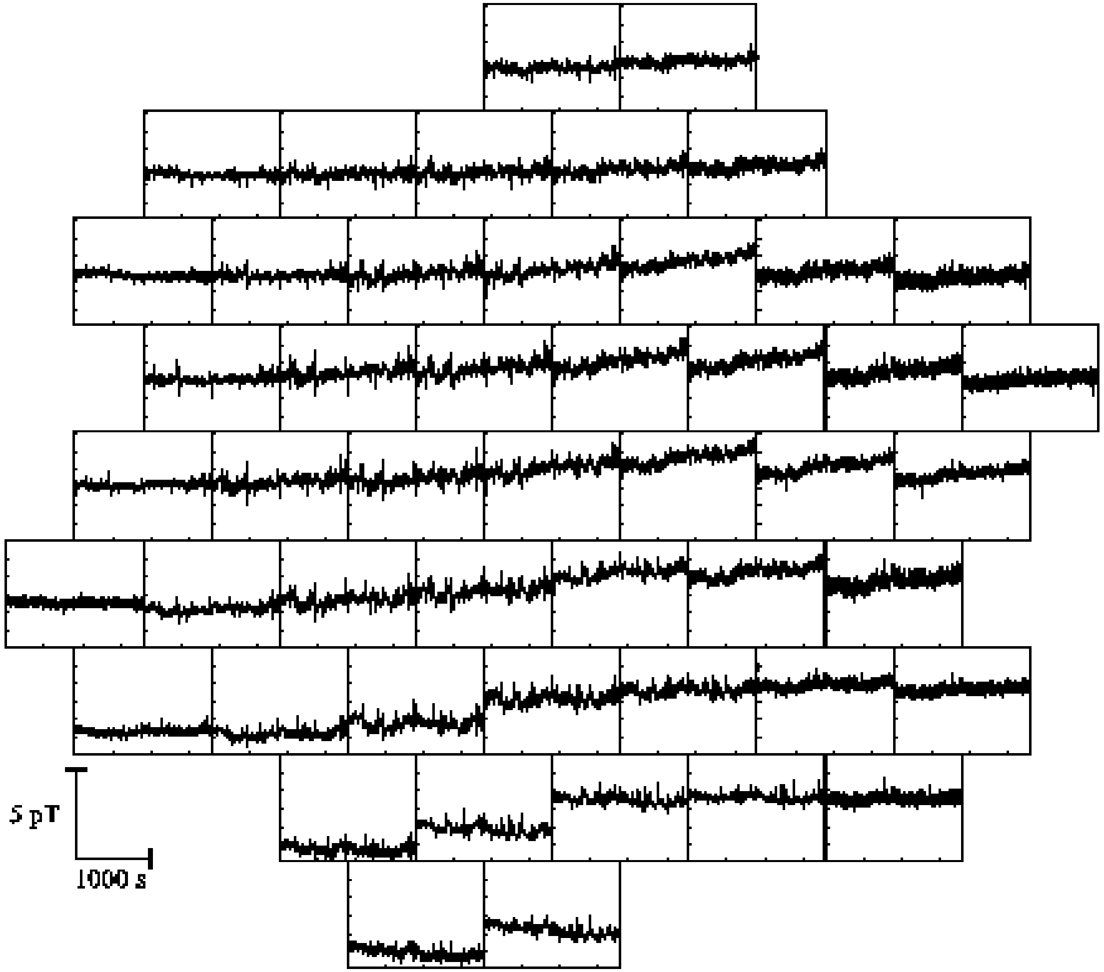
data: reconstructed DC magnetic fields, sampling with modula-tion frequency 0.4 Hz → 720 points/channel for 30 minutes



-

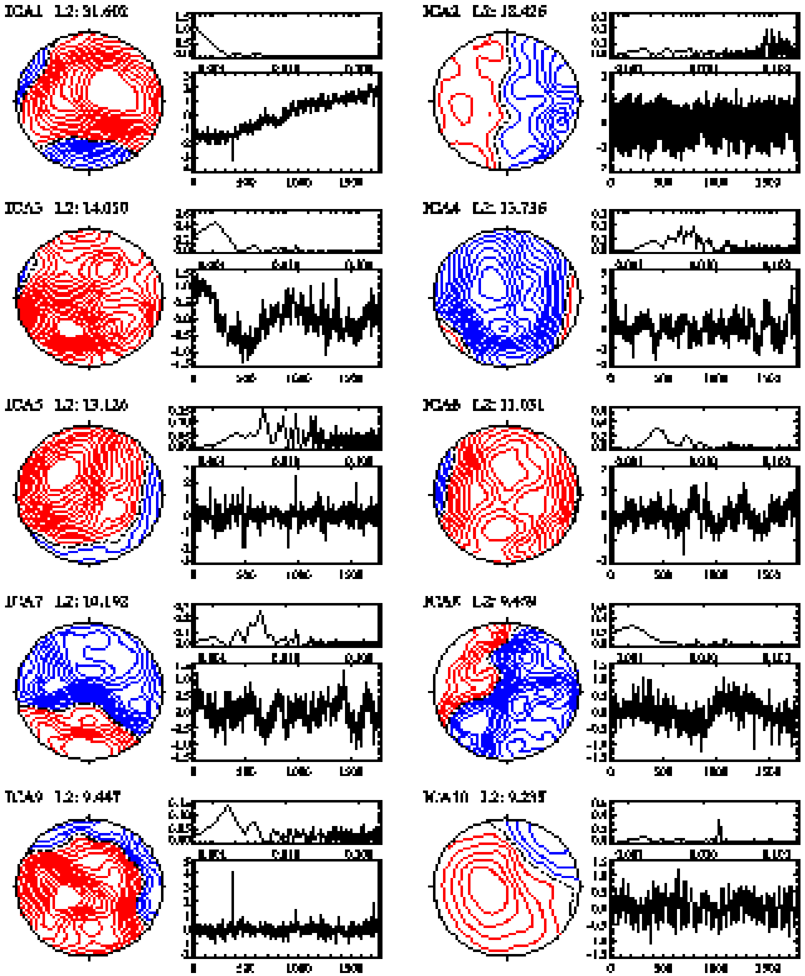
-

Data

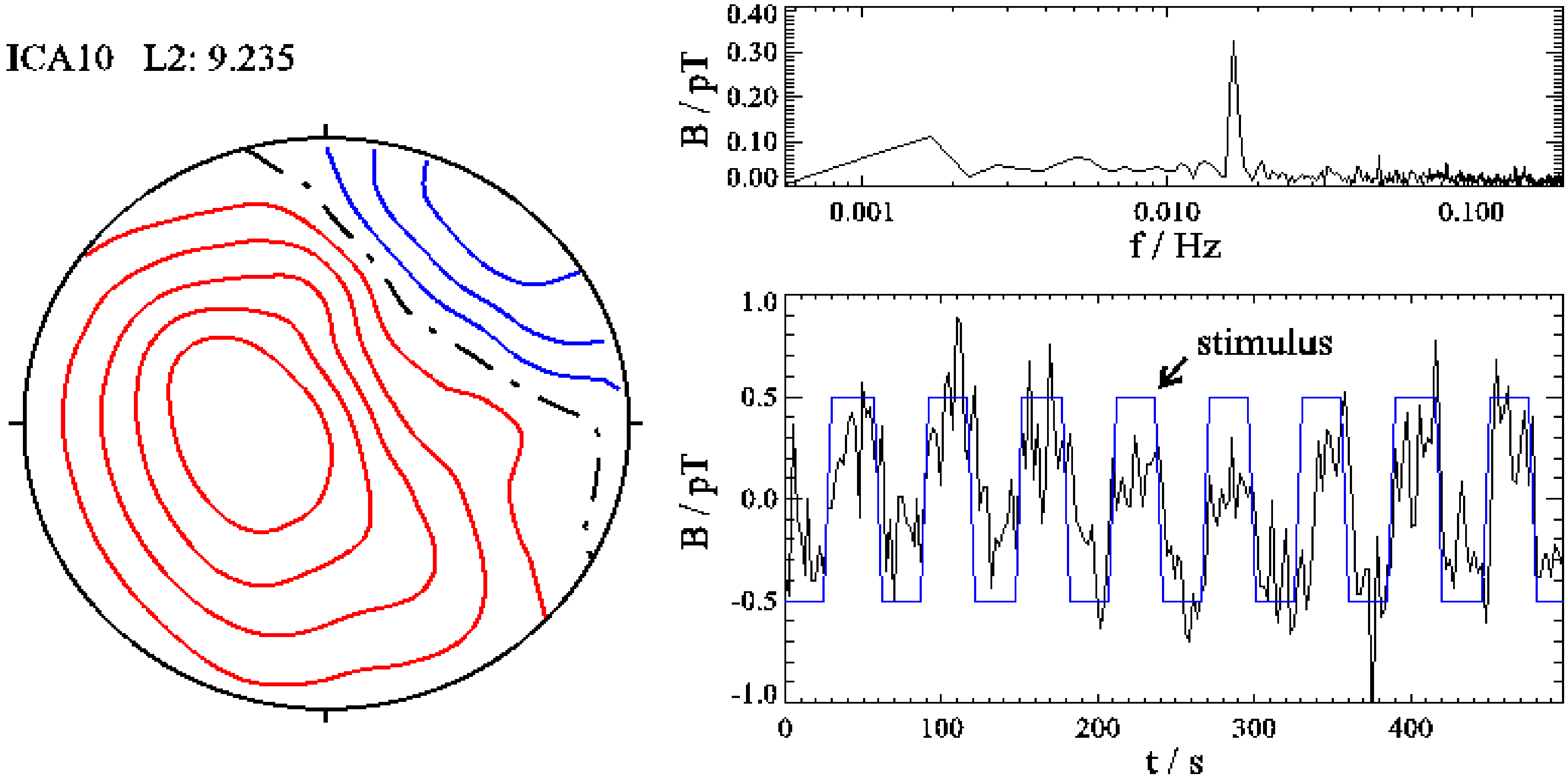


measured data ordered according to sensor position.

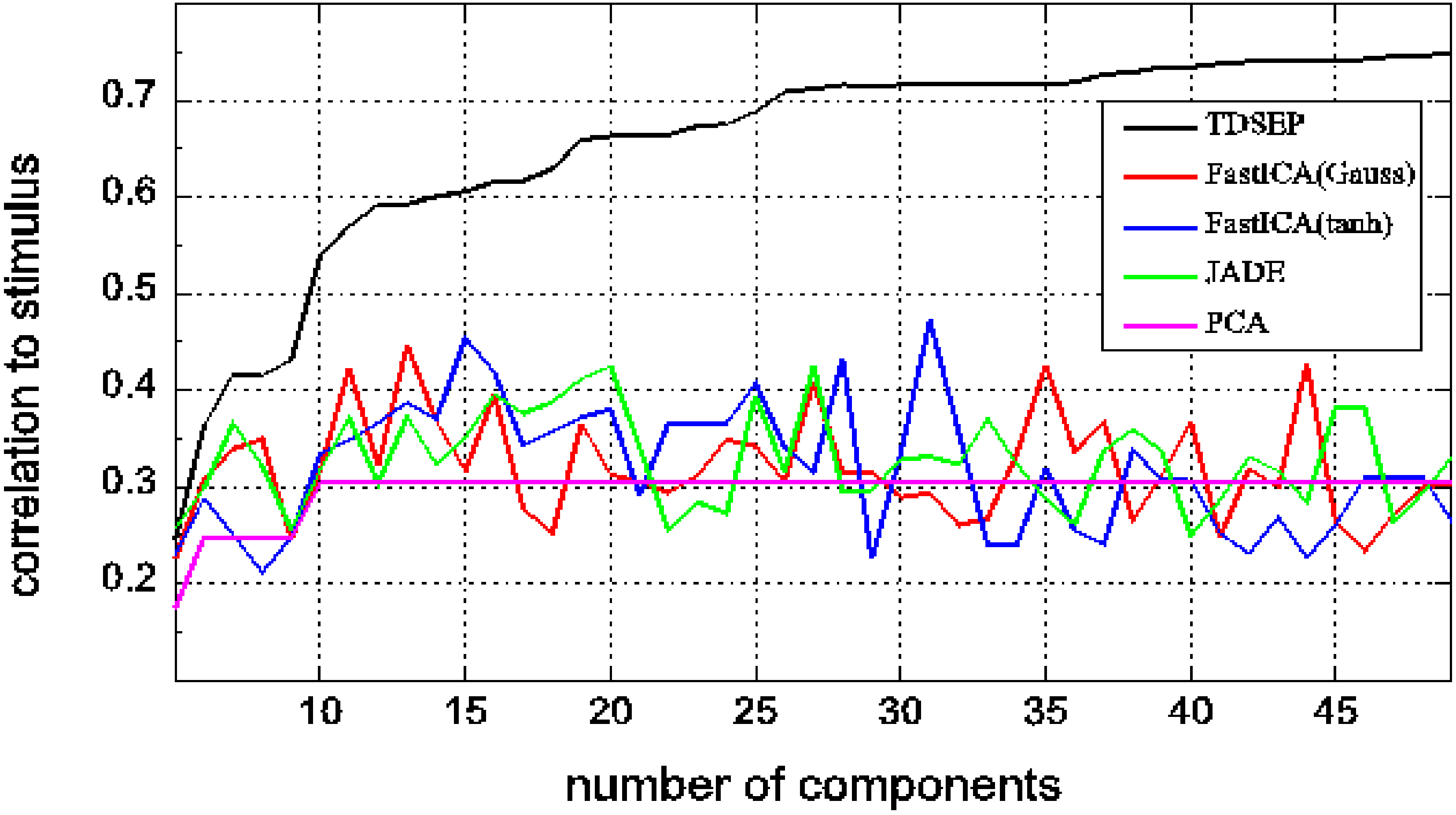
Several ICA Components



Component 10



Comparing three Algorithms



Overview

Overview IDA activities

Introduction to ICA and blind source separation

Reliability and BSS

Non Gaussian Component Analysis

Overview

Overview IDA activities

Introduction to ICA and blind source separation

Reliability and BSS

Non Gaussian Component Analysis

Overview

Overview IDA activities

Introduction to ICA and blind source separation

Reliability and BSS

Non Gaussian Component Analysis

Synchronization and BSS

Conclusions

- efficient and elegant algorithms for linear blind source separation
- reliability analysis of ICA/BSS
- non-Gaussian Component Analysis
- nonlinear BSS (kernel-based methods, Gaussianization, …)
- applications