

# Matrix Factorisation And Its Application In Brain Imaging

Ingo R. Keck  
Centre for Advanced Data Analytics Research  
Ingo.Keck@dit.ie

AIRC/School of Computing/CeADAR Research Seminar

24.09.2013

## Brain Imaging

- fMRI

- fMRI Characteristics

- Study Design

- GLM

- Blind Source Separation

## Matrix Factorisation

- Mixing

- PCA

- ICA

- NMF

## Application fMRI

- spatial ICA

- temporal ICA

- The Noise Problem

- Component Classification

## Application: Resting State

### Experiment

- Resting State

- Networks

## NMF Example

- Example: Retinotopic mapping

## GLM vs ICA

- Overview

- Analysis using SPM

- Analysis using FastICA

## Summary

## Brain Imaging

- fMRI

- fMRI Characteristics

- Study Design

- GLM

- Blind Source Separation

## Matrix Factorisation

- Mixing

- PCA

- ICA

- NMF

## Application fMRI

- spatial ICA

- temporal ICA

- The Noise Problem

- Component Classification

## Application: Resting State

### Experiment

- Resting State

- Networks

## NMF Example

- Example: Retinotopic  
mapping

## GLM vs ICA

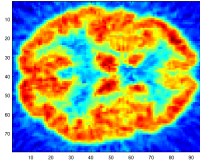
- Overview

- Analysis using SPM

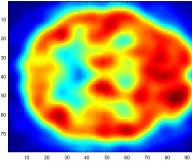
- Analysis using FastICA

## Summary

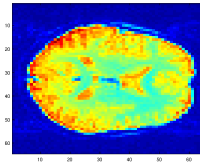
# All Around Brain Activity



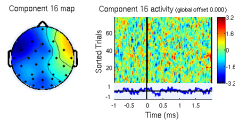
PET



SPECT



fMRI



EEG

# Functional Brain Imaging

METHOD	RESOLUTION	
	TEMPORAL	SPATIAL
MEG	1 ms	5 mm
EEG	1 ms	10–15 mm
fMRI	3–5 s	1.0–1.5 mm
PET	45 s	4 mm

fMRI: good resolution & non-invasive

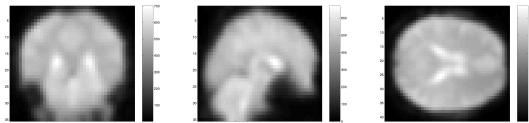
# MRI



## T1 image



# functional Magnetic Resonance Imaging



- ▶ first published by Ogawa et al (1990) and Kwong et al (1992)
- ▶ based on magnetic resonance imaging (MRI)
- ▶ basic unit: 3d volume element “voxel”
- ▶ measures blood flow with haemoglobin as marker (BOLD)



# Activation and Blood Flow

Measurement:

- ▶ neural activity leads to local increase in blood flow in the brain

background:

- ▶ CO<sub>2</sub> increase capillary bed
- ▶ neuronal or → synaptic activity?
- ▶ total brain pressure is kept constant (very complicated system, still not completely understood)

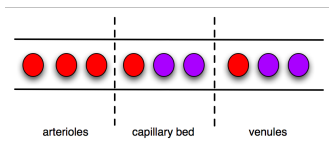
idea:

- ▶ blood flow related to brain activation?
- ▶ we can measure it with marker substances (invasive or natural)

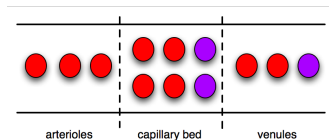
# Blood Oxygenation Level Dependency (BOLD)

Measurement:

- ▶ more local activation in the brain  
→ oxygenation of the blood changes in that region



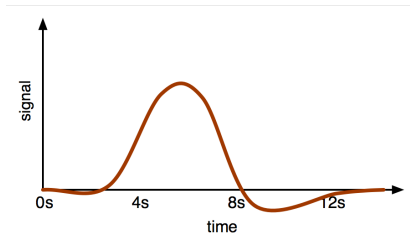
basic activity



increased activity

⇒ lower concentration of deoxyhemoglobine in active regions

# Hemoglobine response function



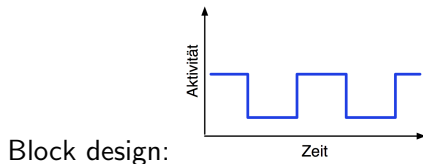
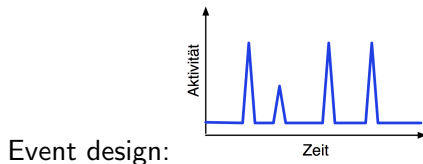
problems:

- ▶ different persons have different HRF
- ▶ depends on position in the brain
- ▶ depends on history of the experiment
- ▶ coffee?

# Compare

[illegible]

Problem: The brain is always active!  $\Rightarrow$  Define basic activity



- ▶ signal of one voxel:

$$\vec{x} = \mathbf{Y}\vec{b} + \vec{e}$$

- ▶  $\mathbf{Y}$ : combination of filter- and activation-functions
- ▶ find regressions coefficients  $\vec{b}$  that minimise  $\vec{e}$

Whole Brain:

$$\mathbf{X} = \mathbf{Y}\beta + \epsilon$$

Basic model: *Which regions correlate with the estimated activation function (based on the stimulus protocol)?*

- ▶ signal of one voxel:

$$\vec{x} = \mathbf{Y}\vec{b} + \vec{e}$$

- ▶  $\mathbf{Y}$ : combination of filter- and activation-functions
- ▶ find regressions coefficients  $\vec{b}$  that minimise  $\vec{e}$

Whole Brain:

$$\mathbf{X} = \mathbf{Y}\beta + \epsilon$$

Basic model: *Which regions correlate with the estimated activation function (based on the stimulus protocol)?*

How well can one estimate the activation in the brain?

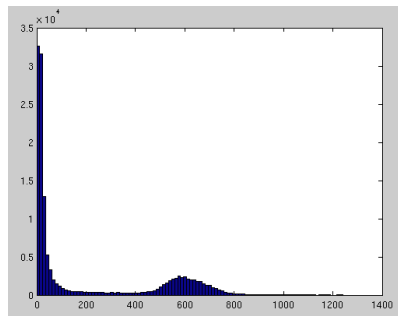
- ▶ direct stimulus: OK
- ▶ brain is involved (aka "thinking"): very difficult!

# Alternatives?

⇒ Exploratory Data Analysis



# Characteristics of fMRI data



- ▶ linear superposition of activation
- ▶ instantaneous mixing
- ▶ stationary (spatial / temporal)
- ▶ non-negative signals
- ▶ independent sources of brain activity

# Blind Source Separation

## Model free analysis: Blind Source Separation (BSS)

- ▶ Signals are orthogonal, variance differs:

Principal Component Analysis

- ▶ Signals are statistically independent:

Independent Component Analysis

- ▶ Signals are non-negative:

Non-negative Matrix Factorisation

# Blind Source Separation

Model free analysis: Blind Source Separation (BSS)

- ▶ Signals are orthogonal, variance differs:

Principal Component Analysis

- ▶ Signals are statistically independent:

Independent Component Analysis

- ▶ Signals are non-negative:

Non-negative Matrix Factorisation

⇒ Explorative, no model testing

## Brain Imaging

- fMRI

- fMRI Characteristics

- Study Design

- GLM

- Blind Source Separation

## Matrix Factorisation

- Mixing

- PCA

- ICA

- NMF

## Application fMRI

- spatial ICA

- temporal ICA

- The Noise Problem

- Component Classification

## Application: Resting State

### Experiment

- Resting State

- Networks

## NMF Example

- Example: Retinotopic mapping

## GLM vs ICA

- Overview

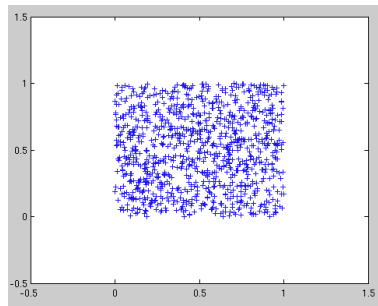
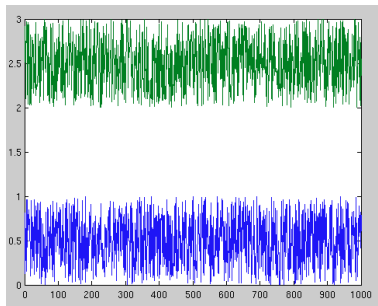
- Analysis using SPM

- Analysis using FastICA

## Summary

# Random Variables

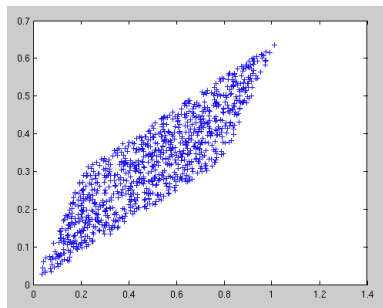
2 random univariant variables  $s_1$  and  $s_2$ :



$$\mathbf{A} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = \mathbf{A}\mathbf{S} =: \mathbf{X}$$

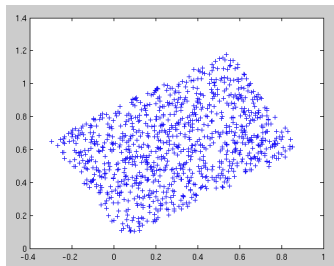
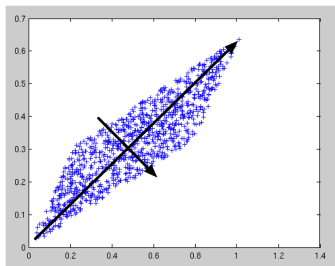
example:

$$\mathbf{A} = \begin{pmatrix} 0.7926 & 0.2235 \\ 0.3290 & 0.3124 \end{pmatrix}$$



Principal Component Analysis searches for orthogonal subspaces of maximal variance:

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}$$



Used for whitening (right) and de-noising.

# PCA - Howto (Actually, it's a SVD)

If  $\mathbf{X}$  is a positive-definite matrix with size  $(m, n)$  and rank  $r$ , then there exists an eigenvector decomposition of the

- ▶ Covariance matrix  $\mathbf{X}\mathbf{X}^T$  with the eigenvector matrix  $\mathbf{U}$
- ▶ Kernel matrix  $\mathbf{X}^T\mathbf{X}$  with the eigenvector matrix  $\mathbf{V}$
- ▶ who share the same  $r$  eigenvalues  $\lambda_i$

Factorise  $\mathbf{X}$  then as follows:

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

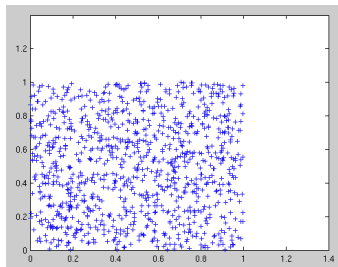
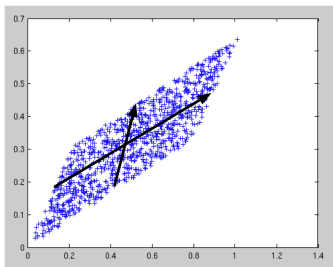
with the roots of the eigenvalues  $\lambda_i$  on the diagonal of  $\mathbf{D}$ .

$\mathbf{U}$  and  $\mathbf{V}$  can be interpreted as rotations, while  $\mathbf{D}$  corresponds to a stretching.



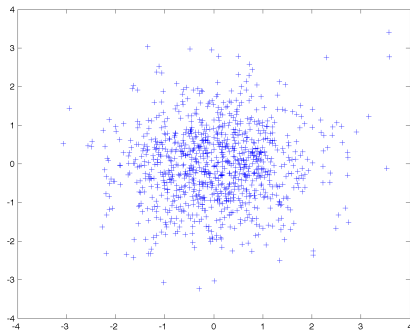
Independent Component Analysis searches for subspaces of maximal independence:

$$\mathbf{X} = \hat{\mathbf{A}}\hat{\mathbf{S}} \approx \mathbf{A}\mathbf{S}$$



Independent generators create independent signals!

Independent Component Analysis can not separate isotropic mixtures:



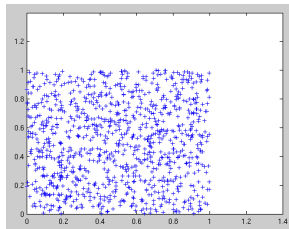
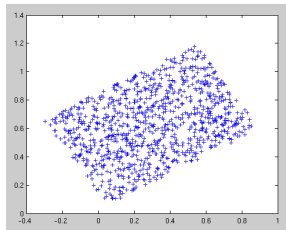
⇒ Gaussian noise can not be separated.

1. Whiten the data with a PCA:

$$\mathbf{W}\mathbf{X} = \mathbf{U}\mathbf{D}^{-1/2}\mathbf{U}^T\mathbf{X}$$

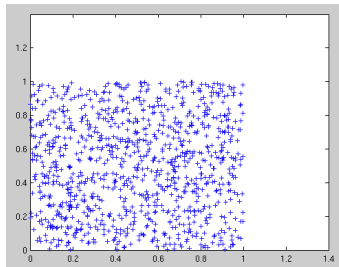
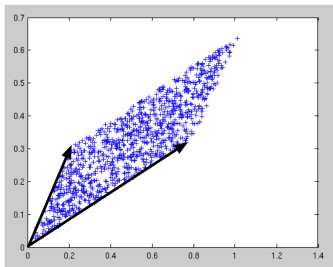
2. Estimate the rotation:

- ▶ Minimise (estimators of) mutual information of de-mixed signals: FastICA
- ▶ Maximise (estimators of) distance to gaussian distribution for de-mixed signals: Infomax, FastICA
- ▶ Geometric approaches: GeolCA



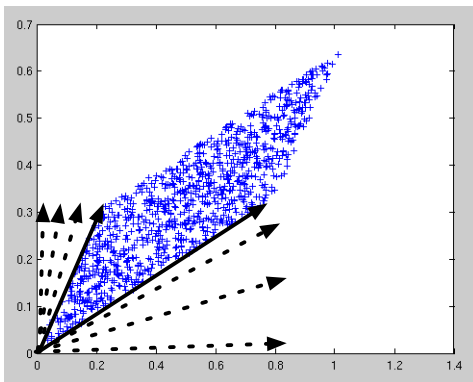
Non-negative Matrix Factorisation searches for non-negative separations:

$$\mathbf{X} = \mathbf{AS} \approx \mathbf{WH}$$



Non-negative Matrix factorisation has an uniqueness problem:

$$\mathbf{X} \approx \mathbf{WH}$$



Estimate the original source matrices:

$$\mathbf{X} \approx \hat{\mathbf{W}}\hat{\mathbf{H}} =: \hat{\mathbf{X}}$$

Basic approach: Find solution that minimises mean squared error (l2-norm) under the constraints of non-negativity:

$$(\mathbf{W}, \mathbf{H}) = \arg \min \left( \sum_{i,j} (x_{ij} - \hat{x}_{ij})^2 \right), \text{ for all } w_{ij}, h_{ij} \geq 0$$

## Brain Imaging

- fMRI

- fMRI Characteristics

- Study Design

- GLM

- Blind Source Separation

## Matrix Factorisation

- Mixing

- PCA

- ICA

- NMF

## Application fMRI

- spatial ICA

- temporal ICA

## The Noise Problem

## Component Classification

## Application: Resting State

## Experiment

- Resting State

- Networks

## NMF Example

- Example: Retinotopic mapping

## GLM vs ICA

- Overview

- Analysis using SPM

- Analysis using FastICA

## Summary

# spatial Independent Component Analysis

slICA applied to fMRI

- ▶ first published by McKeown et al (1998)

Why slICA?

- ▶ functional segregation of the brain → spatial statistical independence
- ▶ “model free”

ICA model:

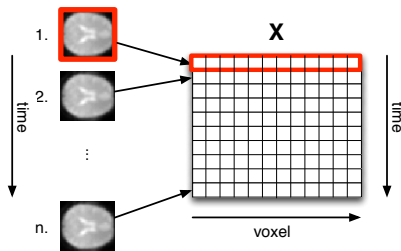
$$\mathbf{X} = \mathbf{AS}$$

- ▶ fMRI-data at one time point as mixture row-vector  $\vec{x}$
- ▶ **S**: independent BOLD-signals of brain activity
- ▶ **A**: matrix of time courses



# Construct Matrix $\mathbf{X}$

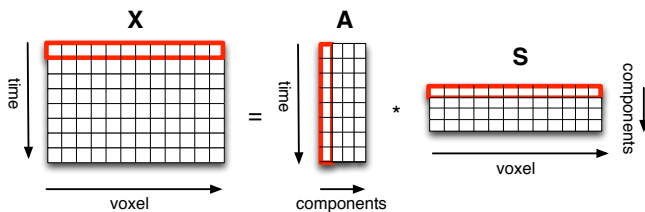
Every volume is rearranged to form one row vector in the data matrix  $\mathbf{X}$ :



$$\mathbf{X} = \mathbf{A}\mathbf{S}$$

# ICA Matrix Model

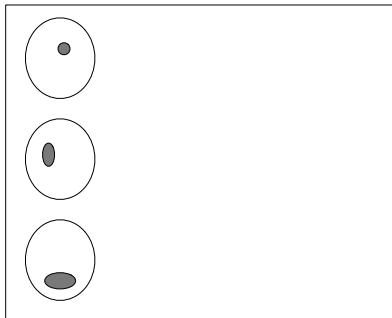
Every row in **X** is constructed by multiplying a column in **A** with the corresponding row in **S**:



$$X = AS$$

# Spatial Mixing Model

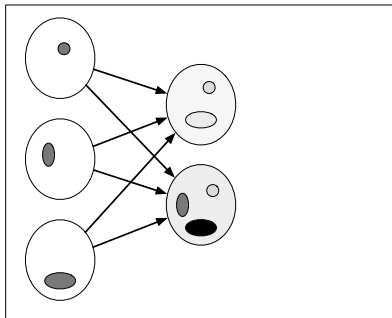
Activations in the brain are combinations of independent spatial activation patterns:



$$\mathbf{X} = \mathbf{AS}$$

# Spatial Mixing Model

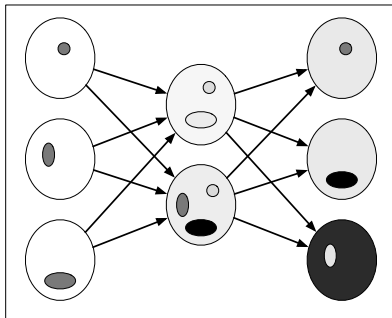
Activations in the brain are combinations of independent spatial activation patterns:



$$\mathbf{X} = \mathbf{AS}$$

# Spatial Mixing Model

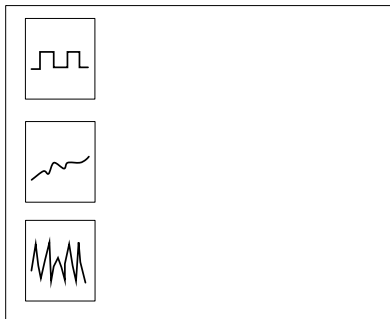
Activations in the brain are combinations of independent spatial activation patterns:



$$\mathbf{X} = \mathbf{AS}$$

# Temporal Mixing Model

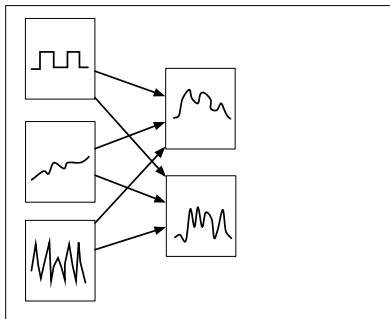
Activations in the brain are combinations of independent temporal activation time courses:



$$\mathbf{X}^T = \mathbf{A}\mathbf{S}$$

# Temporal Mixing Model

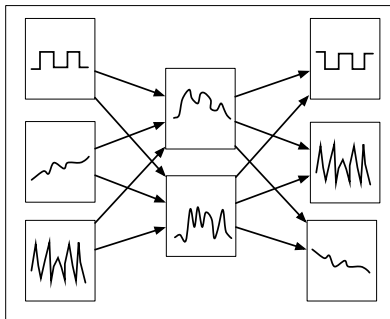
Activations in the brain are combinations of independent temporal activation time courses:



$$\mathbf{X}^T = \mathbf{A}\mathbf{S}$$

# Temporal Mixing Model

Activations in the brain are combinations of independent temporal activation time courses:



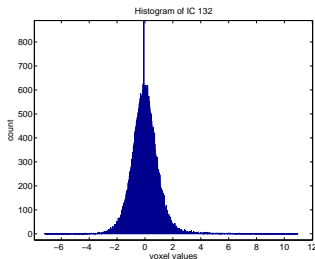
$$\mathbf{X}^T = \mathbf{A}\mathbf{S}$$



# The noise problem

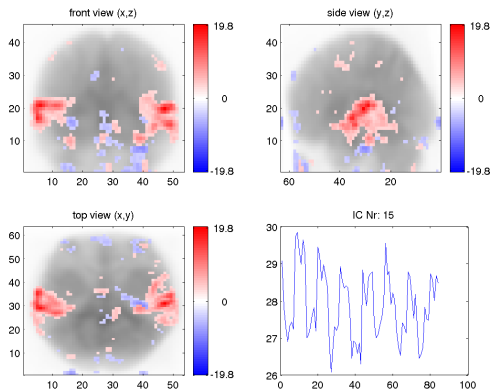
- ▶ fMRI data has low signal-to-noise ratio
- ▶  $\Rightarrow$  noise can not be neglected
- ▶ overcomplete/undetermined problem

- ▶ ICA algorithms robust against noise
- ▶ independent component superimposed on noise background



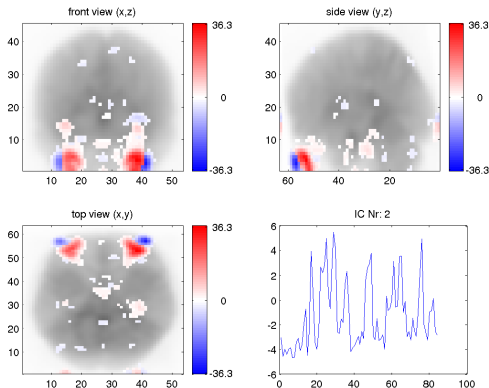
# Component Classification: task related component

From an auditory experiment:

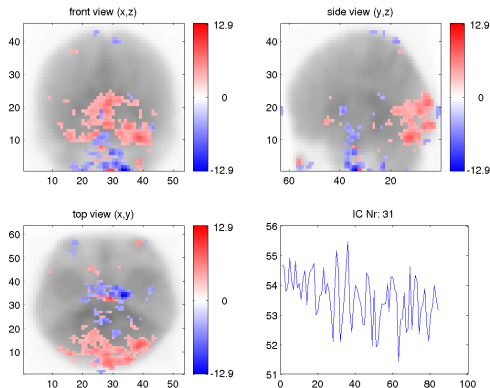


# (transient) movement artefact

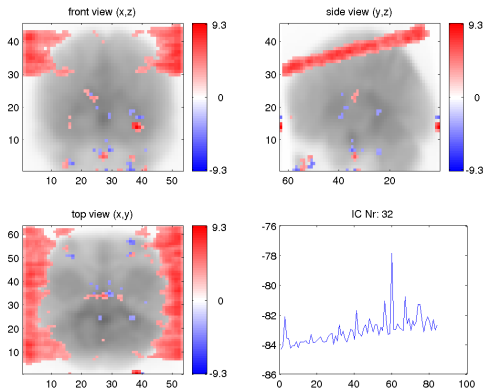
From an auditory experiment:



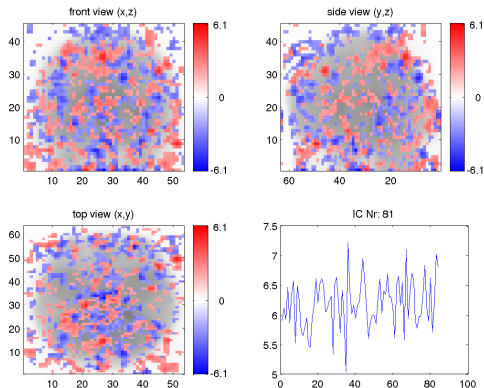
From an auditory experiment:



From an auditory experiment:



From an auditory experiment:



## Brain Imaging

fMRI

fMRI Characteristics

Study Design

GLM

Blind Source Separation

## Matrix Factorisation

Mixing

PCA

ICA

NMF

## Application fMRI

spatial ICA

temporal ICA

The Noise Problem

Component Classification

## Application: Resting State

### Experiment

Resting State

Networks

## NMF Example

Example: Retinotopic  
mapping

## GLM vs ICA

Overview

Analysis using SPM

Analysis using FastICA

## Summary

# Resting State

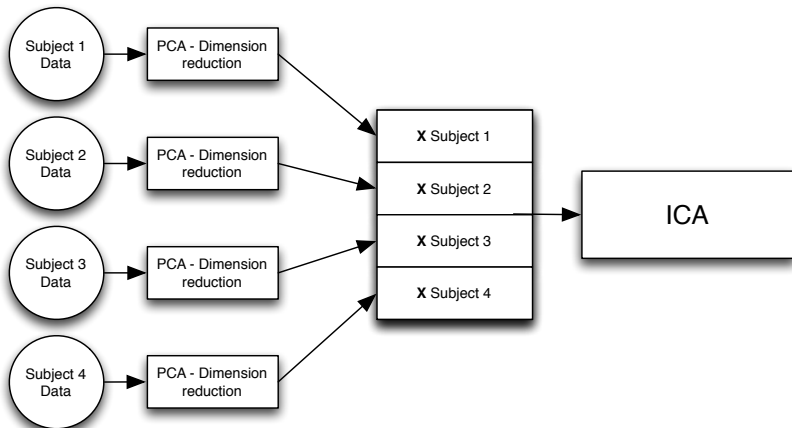
In early fMRI experiments increased brain activity was found during rest events (doing nothing).

Resting state experiment:

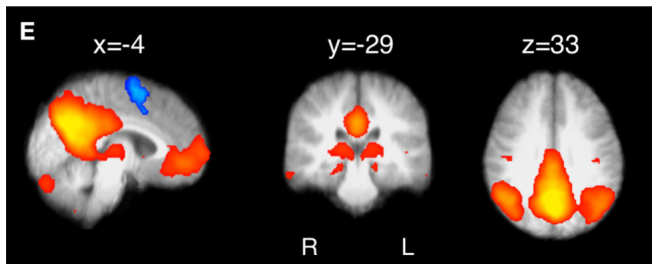
- ▶ subject is instructed to close the eyes and think nothing special
- ▶ experiment length between 5 and 30 minutes (sleep!)
- ▶ no task – no baseline
- ▶ GLM can not be used
- ▶ explorative data analysis like ICA is necessary!



# Group ICA

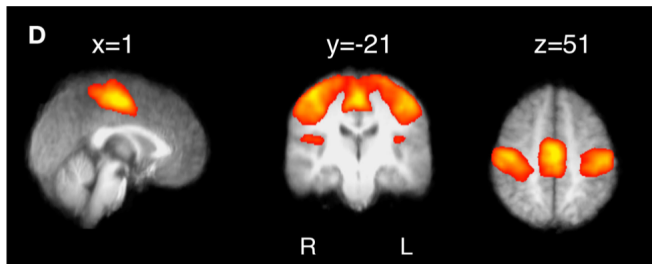


# Default Mode Network



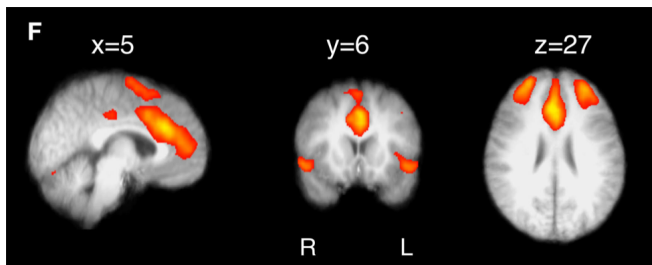
Cole et al, Advances and pitfalls in the analysis and interpretation of resting-state FMRI data, *Frontiers in Systems Neuroscience*, doi: 10.3389/fnsys.2010.00008

# Somatomotor Network



Cole et al, Advances and pitfalls in the analysis and interpretation of resting-state FMRI data, *Frontiers in Systems Neuroscience*, doi: 10.3389/fnsys.2010.00008

# Executive Control and Salience Processing Network



Cole et al, Advances and pitfalls in the analysis and interpretation of resting-state fMRI data, *Frontiers in Systems Neuroscience*, doi: 10.3389/fnsys.2010.00008

## Brain Imaging

fMRI

fMRI Characteristics

Study Design

GLM

Blind Source Separation

## Matrix Factorisation

Mixing

PCA

ICA

NMF

## Application fMRI

spatial ICA

temporal ICA

The Noise Problem

Component Classification

Application: Resting State

Experiment

Resting State

Networks

## NMF Example

Example: Retinotopic  
mapping

## GLM vs ICA

Overview

Analysis using SPM

Analysis using FastICA

## Summary

NMF was re-invented various times:

- ▶ 1971: Lawton, Sylvestre, *Self modeling curve resolution*, Technometrics.
- ▶ 1974: Paatero, Tapper, *Positive matrix factorisation*, Environmetrics.
- ▶ 1999: Lee, Seung, *Learning the parts of objects by non-negative matrix factorisation*, Nature.
- ▶ 1999: Parra, Spence, Sajda, Ziehe, Müller, *Unmixing hyperspectral data*, NIPS.
- ▶ 2000: Lee, Seung, *Algorithms for Non-negative Matrix factorisation*. NIPS.

# Non-Negative Matrix factorisation

NMF applied to fMRI

- ▶ first published by Wang et al (2004)

Why NMF?

- ▶ fMRI signals are positive
- ▶ source mixing is positive

NMF model:

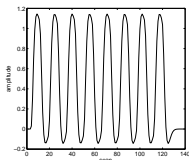
$$\mathbf{X} = \mathbf{WH}$$

- ▶ fMRI-data at one time point as mixture row-vector  $\vec{x}$ , that construct the data matrix  $\mathbf{X}$
- ▶  $\mathbf{H}$ : independent BOLD-signals of brain activity
- ▶  $\mathbf{W}$ : matrix of time courses

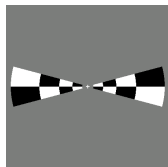
# NMF Example



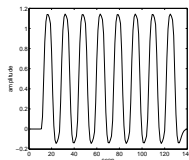
*stimulus design 1*



*onset 1 folded with HRF*



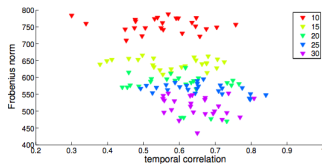
*stimulus design 2*



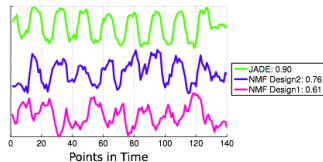
*onset 2 folded with HRF*



## Results:

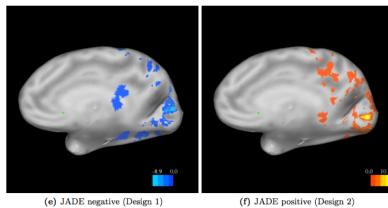
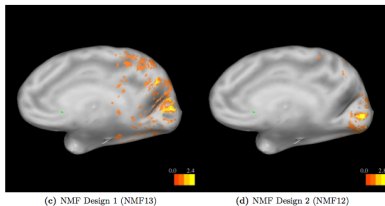


Stochastic results, quality does not depend on number of extracted sources



ICA (JADE) is not able to separate the horizontal and vertical component, NMF can.

# NMF spatial localization



## Brain Imaging

fMRI

fMRI Characteristics

Study Design

GLM

Blind Source Separation

## Matrix Factorisation

Mixing

PCA

ICA

NMF

## Application fMRI

spatial ICA

temporal ICA

The Noise Problem

Component Classification

## Application: Resting State

### Experiment

Resting State

Networks

## NMF Example

Example: Retinotopic  
mapping

## GLM vs ICA

Overview

Analysis using SPM

Analysis using FastICA

## Summary

# Word perception experiment

## Setting:

- ▶ 4 different subjects
- ▶ each subject listens to mutated words
- ▶ 4 classes, based on number of frequency bands (FB1–FB4)
- ▶ only class FB4 perceivable as words
- ▶ Task: press button if the word was understood

# Word perception experiment

## Setting:

- ▶ 4 different subjects
- ▶ each subject listens to mutated words
- ▶ 4 classes, based on number of frequency bands (FB1–FB4)
- ▶ only class FB4 perceivable as words
- ▶ Task: press button if the word was understood

## Expected outcome:

- ▶ activation in areas for word detection
- ▶ perhaps selective activation in auditory cortex for FB1–FB4

# Word perception experiment

## Setting:

- ▶ 4 different subjects
- ▶ each subject listens to mutated words
- ▶ 4 classes, based on number of frequency bands (FB1–FB4)
- ▶ only class FB4 perceivable as words
- ▶ Task: press button if the word was understood

## Expected outcome:

- ▶ activation in areas for word detection
- ▶ perhaps selective activation in auditory cortex for FB1–FB4

## Analysis:

- ▶ single subject analysis

Expected timecourses based on experiment protocol

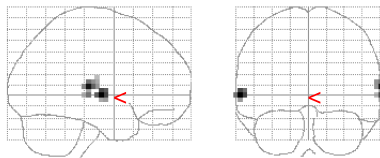
- ▶ no significantly correlated activation for one subject

# Analysis using SPM

Expected timecourses based on experiment protocol

- ▶ no significantly correlated activation for one subject
- ▶ marginal activation for FB4 in two subjects

FB4



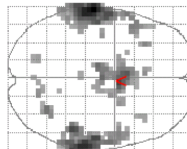
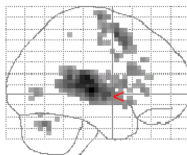
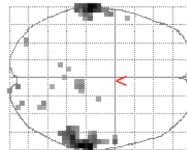
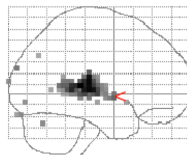


# Analysis using SPM

Expected timecourses based on experiment protocol

- ▶ no significantly correlated activation for one subject
- ▶ marginal activation for FB4 in two subjects
- ▶ one subject showed activations for FB1 to FB4
  - ▶ auditory cortex (FB1 – FB4)
  - ▶ left supplementary motor area (FB4)
  - ▶ cingulate gyrus (FB4)

FB1



FB4

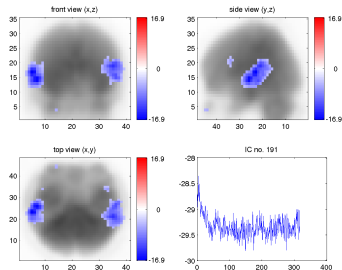
Expected timecourses based on experiment protocol

- ▶ no significantly correlated activation for one subject
- ▶ marginal activation for FB4 in two subjects
- ▶ one subject showed activations for FB1 to FB4
  - ▶ auditory cortex (FB1 – FB4)
  - ▶ left supplementary motor area (FB4)
  - ▶ cingulate gyrus (FB4)

*Where are the activations related to word detection?*

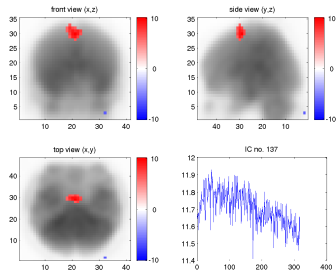
# Spatial ICA with FastICA

- ▶ various components with activation in the auditory cortex in all subjects



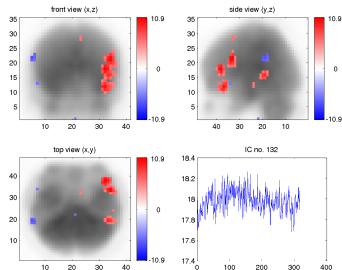
# Spatial ICA with FastICA

- ▶ various components with activation in the auditory cortex in all subjects
- ▶ components with activation in the supplementary motor areas in all subjects



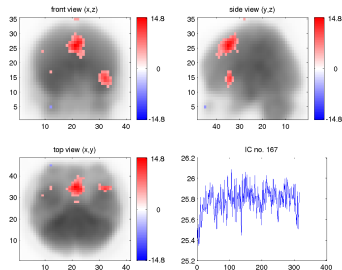
# Spatial ICA with FastICA

- ▶ various components with activation in the auditory cortex in all subjects
- ▶ components with activation in the supplementary motor areas in all subjects
- ▶ components related to word detection network in three subjects

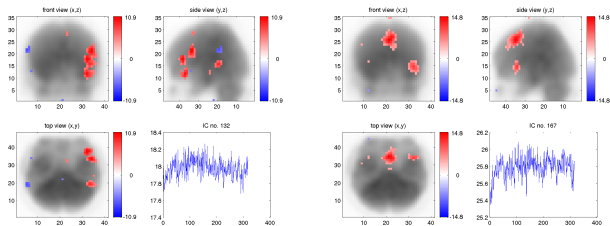


# Spatial ICA with FastICA

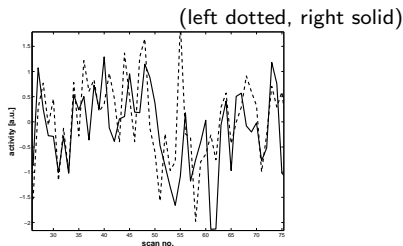
- ▶ various components with activation in the auditory cortex in all subjects
- ▶ components with activation in the supplementary motor areas in all subjects
- ▶ components related to word detection network in three subjects
- ▶ components related to decision finding network in three subjects



# Comparison of functional networks' time courses



- ▶ one subject as example
- ▶ visually correlated time courses
- ▶  $k_{\text{corr}} = 0.36$  due to local baseline- and time-shifts
- ▶ only 0.7 % of fMRI data



- ▶ Explorative data analysis is a powerful tool for brain imaging research.
- ▶ PCA generic pre-processing step.
- ▶ ICA reliable and already used in research.
- ▶ NMF can help for statistically dependent signals.