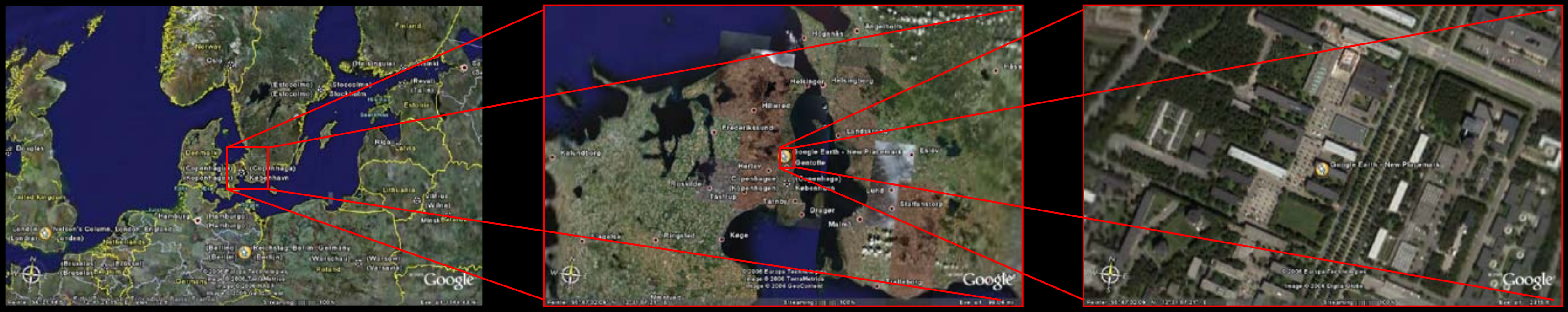
# Shift Invariant Data Decomposition

Morten Mørup
Informatics and Mathematical Modeling
Intelligent Signal Processing
Technical University of Denmark



# Collaborators


## Kristoffer Hougaard Madsen and Lars Kai Hansen and Mikkel N. Schmidt

Informatics and Mathematical Modeling

Intelligent Signal Processing

Technical University of Denmark



**Sidse M. Arnfred, Dr. Med. PhD**

Cognitive Research Unit

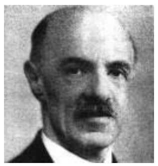
Hvidovre Hospital

University Hospital of Copenhagen

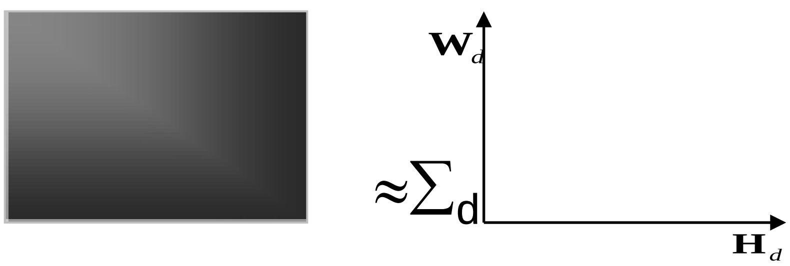
# Outline

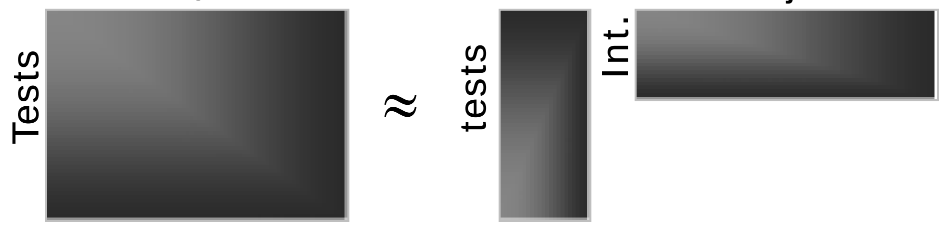
- Shifted Non-negative Matrix Factorization
- Shifted Independent Component analysis
- Generalization to tensors (i.e., the PARAFAC model)
- Shift Invariant Sparse Coding

# Factor Analysis



Spearman ~1900




$$\mathbf{V} \approx \mathbf{W}\mathbf{H}$$

$$\mathbf{V}^{\text{tests x subjects}} \approx \mathbf{W}^{\text{tests x intelligences}}\mathbf{H}^{\text{intelligencesxsubject}}$$

Independent Component Analysis (ICA)
rows of **H** statistically independent

(P. Common, Bell & Sejnowski ~1995)

Non-negative Matrix Factorization (NMF):

$$\mathbf{V}_{n,m}, \mathbf{W}_{n,d}, \mathbf{H}_{d,m} \geq 0$$

(~1970 Lawson, ~1995 Paatero, ~2000 Lee & Seung)

# Multiplicative updates

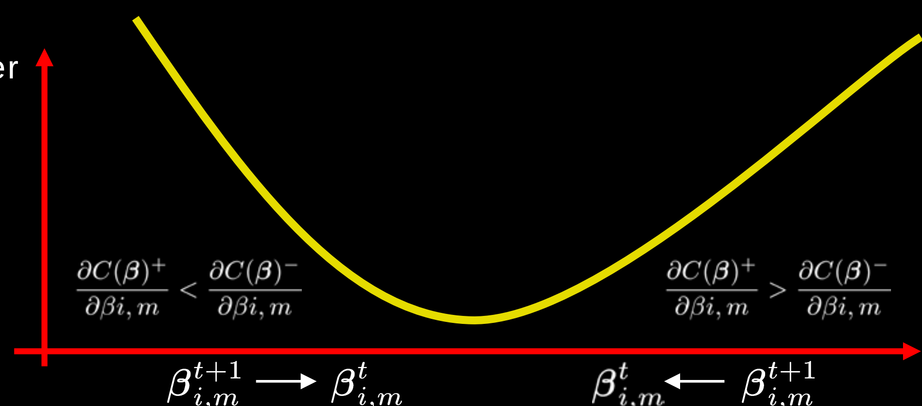
$$\frac{\partial C(\boldsymbol{\beta})}{\partial \beta i,m} = \frac{\partial C(\boldsymbol{\beta})^{+}}{\partial \beta i,m} - \frac{\partial C(\boldsymbol{\beta})^{-}}{\partial \beta i,m}$$

$$\boldsymbol{\beta}_{i,m} \quad = \quad \boldsymbol{\beta}_{i,m} - \mu_{i,m}\frac{\partial C(\boldsymbol{\beta})}{\partial \beta i,m}, \mu_{i,m} = \frac{\boldsymbol{\beta}_{i,m}}{\frac{\partial C(\boldsymbol{\beta})^{+}}{\partial \beta i,m}}$$

$$\boldsymbol{\beta}_{i,m} \quad = \quad \boldsymbol{\beta}_{i,m} - \frac{\boldsymbol{\beta}_{i,m}}{\frac{\partial C(\boldsymbol{\beta})^{+}}{\partial \beta i,m}}\left(\frac{\partial C(\boldsymbol{\beta})^{+}}{\partial \beta i,m} - \frac{\partial C(\boldsymbol{\beta})^{-}}{\partial \beta i,m}\right) = \boldsymbol{\beta}_{i,m}\frac{\frac{\partial C(\boldsymbol{\beta})^{-}}{\partial \beta i,m}}{\frac{\partial C(\boldsymbol{\beta})^{+}}{\partial \beta i,m}}$$

Step size parameter

$$\boldsymbol{\beta}_{i,m}^{t+1} \leftarrow \boldsymbol{\beta}_{i,m}^{t}\left(\frac{\frac{\partial C(\boldsymbol{\beta})^{-}}{\partial \boldsymbol{\beta}_{i,m}^{t}}}{\frac{\partial C(\boldsymbol{\beta})^{+}}{\partial \boldsymbol{\beta}_{i,m}^{t}}}\right)^{\alpha}$$



# Non-negative matrix factorization

$$\mathbf{V}_{i,j} \geq 0 \quad , \quad \mathbf{W}_{i,d} \geq 0 \quad \textit{and} \quad \mathbf{H}_{d,j} \geq 0$$

$$C_{LS} = \frac{1}{2}\|\mathbf{V} - \mathbf{W}\mathbf{H}\|_F^2 = \frac{1}{2}\sum_{i,j}(\mathbf{V}_{i,j} - (\mathbf{W}\mathbf{H})_{i,j})^2$$

$$\mathbf{W}_{i,d} \leftarrow \mathbf{W}_{i,d}\frac{(\mathbf{V}\mathbf{H}^{\mathrm{T}})_{i,d}}{(\mathbf{W}\mathbf{H}\mathbf{H}^{\mathrm{T}})_{i,d}}$$

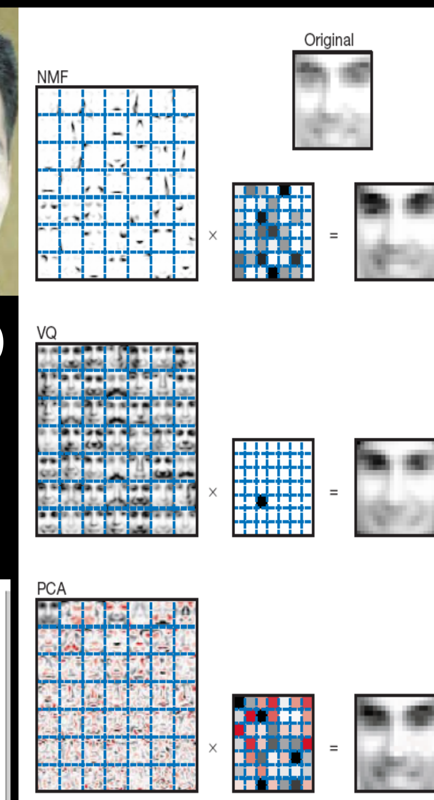
$$\mathbf{H}_{d,j} \leftarrow \mathbf{H}_{d,j}\frac{(\mathbf{W}^{\mathrm{T}}\mathbf{V})_{d,j}}{(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{H})_{d,j}}$$

$$C_{KL} = \sum_{i,j}\mathbf{V}_{i,j}\log\frac{\mathbf{V}_{i,j}}{(\mathbf{W}\mathbf{H})_{i,j}} - \mathbf{V}_{i,j} + (\mathbf{W}\mathbf{H})_{i,j}$$

$$\mathbf{W}_{i,d} \leftarrow \mathbf{W}_{i,d}\frac{\sum_j \frac{\mathbf{V}_{i,j}}{(\mathbf{W}\mathbf{H})_{i,j}}\mathbf{H}_{d,j}}{\sum_j \mathbf{H}_{d,j}}$$

$$\mathbf{H}_{d,j} \leftarrow \mathbf{H}_{d,j}\frac{\sum_i \mathbf{W}_{i,d}\frac{\mathbf{V}_{i,j}}{(\mathbf{W}\mathbf{H})_{i,j}}}{\sum_i \mathbf{W}_{i,d}}$$

(Lee & Seung - 2001)




**Learning the parts of objects by non-negative matrix factorization**

Daniel D. Lee* & H. Sebastian Seung*†

* Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA
† Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Is perception of the whole based on perception of its parts? There

(Some other approaches:
Active Set, projected gradient,
barrier functions, exponentiation)

NMF gives Part based representation
(Lee & Seung – Nature 1999)

# Maximum likelihood (ML) ICA approach

Equivalently we derive two step procedure

**[A',S',H'] = SVD(X)** → Shift Invariant Subspace Analysis (SISA)

**W' = A'S'**

Notice decomposition then ambiguous since

$$\mathbf{WH} = (\mathbf{W'Q}^{-1})(\mathbf{QH'}) = \mathbf{W'H'}$$

Thus ICA forms objective for ambiguity **Q** that minimizes: → Shifted Independent Component Analysis (SICA)

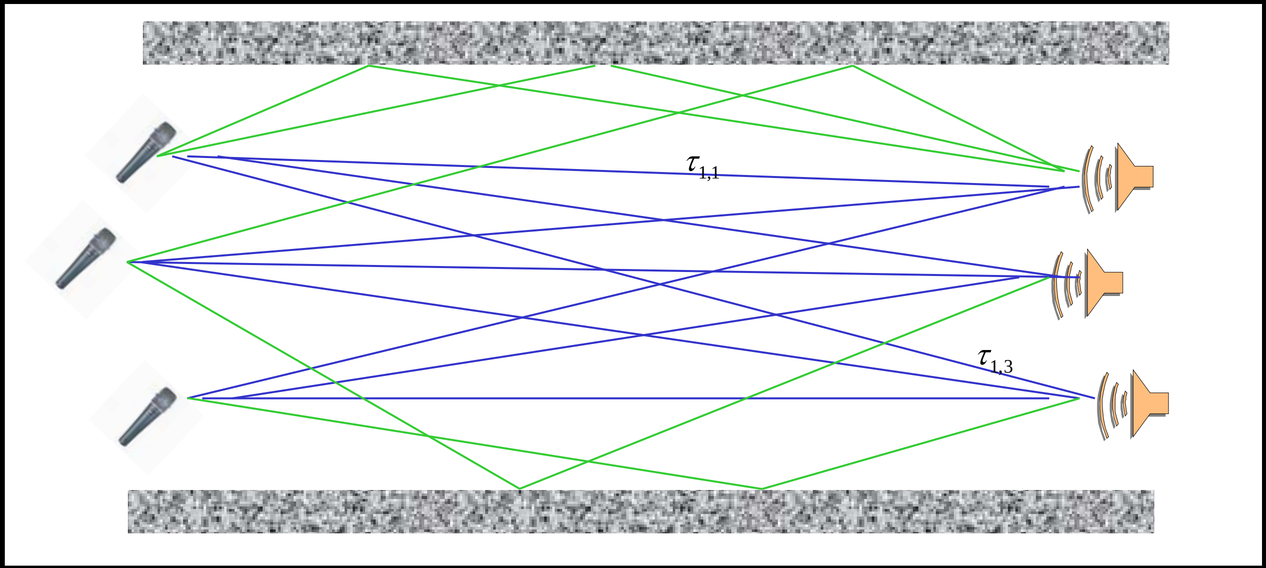
ICA-ML

$$p(\mathbf{H}|\mathbf{Q}) = \prod_m p(\mathbf{H}_m|\mathbf{Q}) = \prod_m |det(\mathbf{Q})| p(\mathbf{Q}\mathbf{H}_m)$$

Assume Independence

Change of variable principle

# The shift problem




Convolutive ICA/NMF
(echo effects, Smaragdis 2003)

Shifted ICA/NMF
(One specific delay between each sensor and source)

$$\mathbf{V}_{n,m} = \sum_{d,\tau} \mathbf{W}^{\tau}_{n,d} \mathbf{H}_{d,m-\tau} + \mathbf{E}_{n,m}$$

$$\mathbf{V}_{n,m} = \sum_{d} \mathbf{W}_{n,d} \mathbf{H}_{d,m-\boldsymbol{\tau}_{n,d}} + \mathbf{E}_{n,m}$$

# History of shift


- Bell & Sejnowski 1995
  (Sketched how to handle time delays in networks)
- Torkkola 1996
  (Further developed Bell and Sejnowski's work)

- Emile & Comon 1998
  (delay in model based on equally mixed sources formed by moving averages)

- Hong and Harshman 2003 – shifted factor analysis
  (a procedure based on exhaustive search over integer shifts – model conjectured unique)

- Yeredor 2003 (Solved the ICA model with shifts by joint diagonalization (sources=sensors) of the source cross spectra based on the AC-DC algorithm with non-integer shifts for the 2x2 system)

- Yeredor 2005 (extension to complex signals)

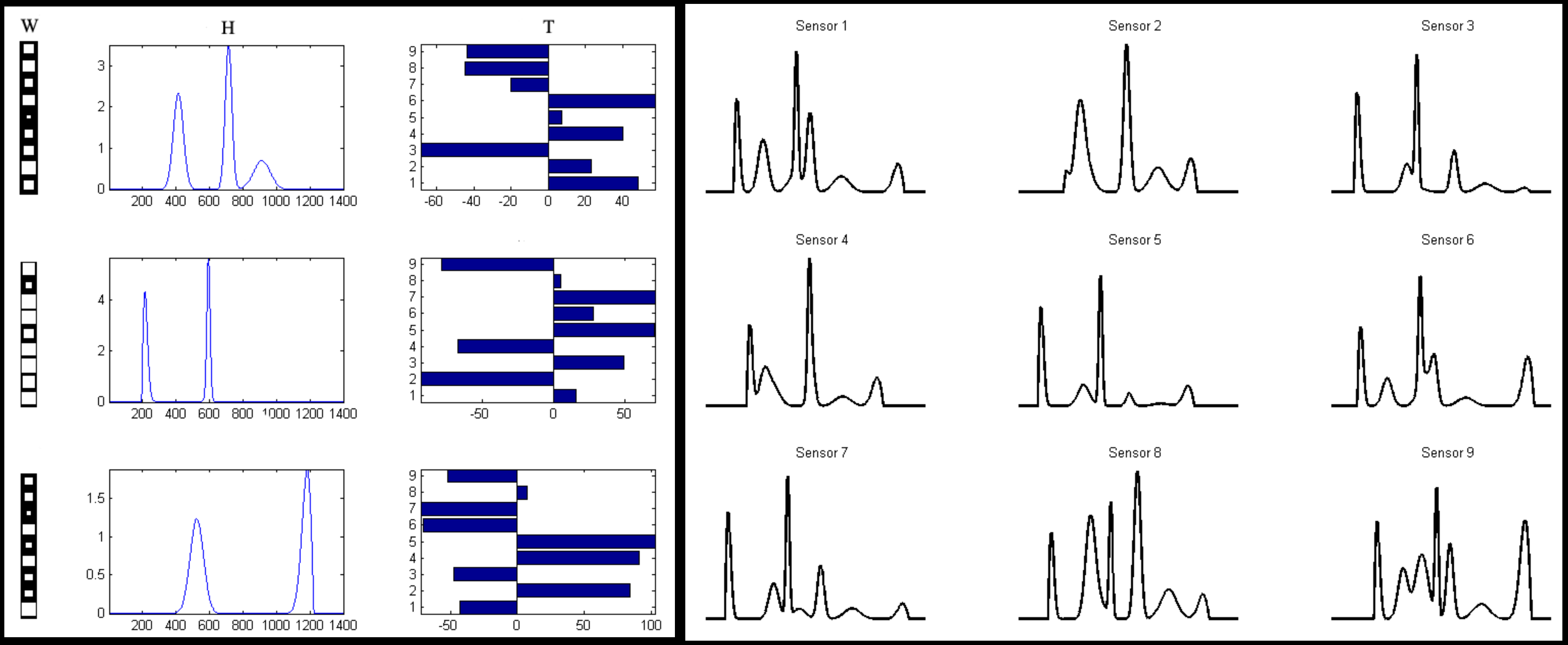
# Why shifted ICA/NMF

Causes of shifts for instance

Doppler effect, Time of arrival differences

- Magnetic resonance spectra
  (Du et al, 2005)
- Astronomical spectrometers (red shift)
  (Pauca et al. 2006)
- Fluorescence spectra
  (Gobinet et al. 2004)
- PET imaging
  (Kim et al. 2001, Lee et al., 2001, Bödvarsson et al. 2007)
- Sound recording (delays between source and sensor due to propagation delay)

# Generative model



# Notation and LS-objective

- $\mathbf{U}$ and $\widetilde{\mathbf{U}}$ denotes same matrix in time and frequency domain respectively.
- $\widetilde{\mathbf{U}}^H$ denotes the conjugate transpose of $\widetilde{\mathbf{U}}$
- $\mathbf{U}_{d,m-\tau} \sim \widetilde{\mathbf{U}}_{d,f} e^{-i2\pi\frac{f-1}{M}\tau}$
- $\mathbf{U} \bullet \mathbf{V}$ denotes the direct product, i.e. element-wise multiplication.
- $(e^{-i2\pi\frac{f-1}{M}\boldsymbol{\tau}})_{n,d} = e^{-i2\pi\frac{f-1}{M}\boldsymbol{\tau}_{n,d}}$.
- $\widetilde{\mathbf{U}}^{(f)} = \mathbf{U} \bullet e^{-i2\pi\frac{f-1}{M}\boldsymbol{\tau}}$
- $\mathbf{U}_d$ $d^{th}$ column, $\mathbf{U}_{n,:}$ $n^{\text{th}}$ row and $\mathbf{U}_{n,d}$ a given element of $\mathbf{U}$.

$$\boxed{\begin{aligned} C_{LS}(W,H) &= \frac{1}{2}\sum_{n,m}(\mathbf{V}_{n,m} - \sum_d \mathbf{W}_{n,d}\mathbf{H}_{d,m-\boldsymbol{\tau}_{n,d}})^2 \\ &= \frac{1}{2M}\|\widetilde{\mathbf{V}}_f - (\mathbf{W} \bullet e^{-i2\pi\frac{f-1}{M}\boldsymbol{\tau}})\widetilde{\mathbf{H}}_f\|_F^2 = \frac{1}{2M}\|\widetilde{\mathbf{V}}_f - \mathbf{W}^{(f)}\widetilde{\mathbf{H}}_f\|_F^2 \end{aligned}}$$

Follows from Parsevals identity with the above notation

# **W** update

Let $\widetilde{\mathbf{H}}^{(n)}_{d,f} = \widetilde{\mathbf{H}}_{d,f} e^{-i2\pi \frac{f-1}{M} \boldsymbol{\tau}_{n,d}}$ denote the delayed version of the source signal $\widetilde{\mathbf{H}}_{d,f}$ to the $n^{\text{th}}$ channel. The shift ICA/NMF model can then be stated as

$$\mathbf{V}_{n,:} = \mathbf{W}_{n,:}\mathbf{H}^{(n)} + \mathbf{E}_{n,:}.$$

This is the regular ICA/NMF problem which can be solved by the least squares ICA/NMF-update

$$\begin{aligned} SISA : \mathbf{W}_{n,:} &= \mathbf{V}_{n,:}/(\mathbf{H}^{(n)}\mathbf{H}^{(n)^T}), \\ NMF : \mathbf{W}_{n,d} &= \mathbf{W}_{n,d}\frac{\mathbf{V}_{n,:}\mathbf{H}^{(n)^T}_{d,:}}{\mathbf{W}_{n,:}\mathbf{H}^{(n)}\mathbf{H}^{(n)^T}_{d,:}}. \end{aligned}$$

# H update

$$C_{LS} = \frac{1}{2M}\|\widetilde{\mathbf{V}}_f - \widetilde{\mathbf{W}}^{(f)}\widetilde{\mathbf{H}}_f\|_F^2$$

$$\mathbf{G}_f = \frac{\partial C_{LS}}{\partial \widetilde{\mathbf{H}}_f} = -\frac{1}{M}\widetilde{\mathbf{W}}^{(f)^H}(\widetilde{\mathbf{X}}_f - \widetilde{\mathbf{W}}^{(f)}\widetilde{\mathbf{H}}_f)$$

## ShiftNMF

$$\begin{aligned}\widetilde{\mathbf{G}}_f^+ &= \frac{1}{M}\widetilde{\mathbf{W}}^{(f)^H}\widetilde{\mathbf{W}}^{(f)}\widetilde{\mathbf{H}}_f \\ \widetilde{\mathbf{G}}_f^- &= \frac{1}{M}\widetilde{\mathbf{W}}^{(f)^H}\widetilde{\mathbf{X}}_f\end{aligned}$$

$$\mathbf{H}_{d,n} = \mathbf{H}_{d,n}\left(\frac{\mathbf{G}_{d,n}^-}{\mathbf{G}_{d,n}^+}\right)^\alpha$$

## SISA

$$\mathbf{G}_f = (\widetilde{\mathbf{W}}^{(f)^\top}\widetilde{\mathbf{W}}^{(f)})\backslash\widetilde{\mathbf{X}}_f$$

# Update of the shifts (τ)

$$C_{LS} = \frac{1}{2M}\sum_f (\widetilde{\mathbf{V}}_f - \ldots (\mathbf{W} \bullet e^{-i2\pi\frac{f-1}{M}\boldsymbol{\tau}})\widetilde{\mathbf{H}}_f)$$

**Warning! Prone to local minima**

wher… …be estimated
by the N…

# Update of shifts (τ) based on Cross-correlation

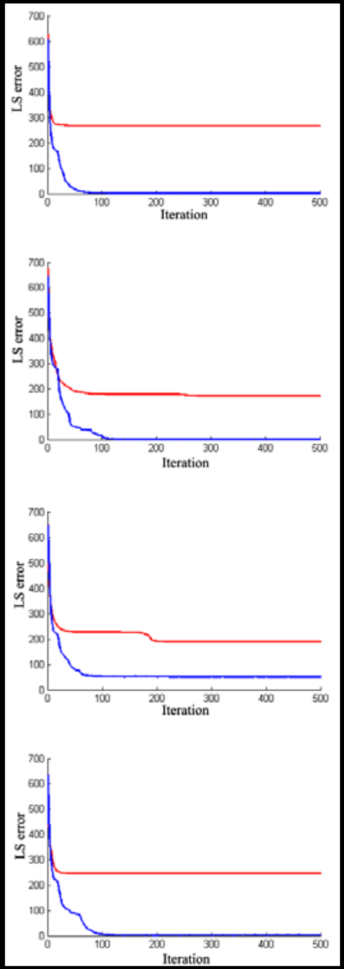
$$
\begin{aligned}
\mathbf{R}_{n,m} &= \mathbf{V}_{n,m} - \sum_{d\neq d'} \mathbf{W}_{n,d}\mathbf{H}_{d,m-\boldsymbol{\tau}_{n,d}} \\
C_{LS} &= \frac{1}{2}\sum_{n,m}(\mathbf{V}_{n,m} - \sum_{d} \mathbf{W}_{n,d}\mathbf{H}_{d,m-\boldsymbol{\tau}_{n,d}})^2 \\
&= \frac{1}{2}\sum_{n,m}(\mathbf{R}_{n.m} - \mathbf{W}_{n,d'}\mathbf{H}_{d',m-\boldsymbol{\tau}_{n,d'}})^2 \\
&= \frac{1}{2}\|\mathbf{R}\|^2 - \sum_{n}\mathbf{W}_{n,d'}\sum_{m}\mathbf{R}_{n,m}\mathbf{H}_{d',m-\boldsymbol{\tau}_{n,d'}} + \frac{1}{2}\|\mathbf{W}\mathbf{H}\|^2
\end{aligned}
$$

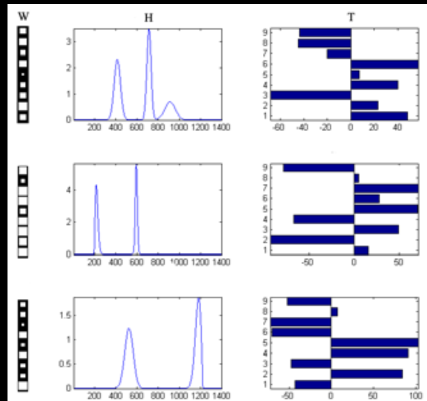
Independent of τ   Cross correlation R and H   Independent of τ

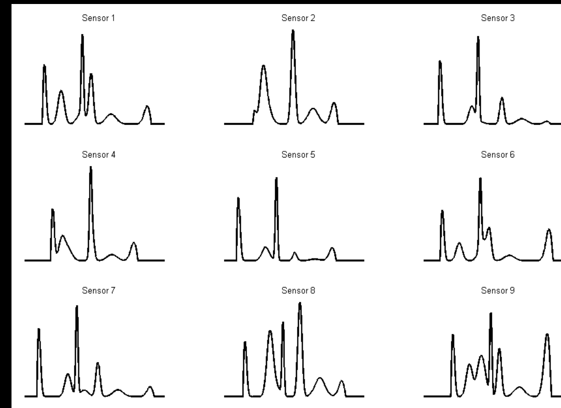
$$
\widetilde{\mathbf{c}}_f = \widetilde{\mathbf{R}}^*_{n,f}\widetilde{\mathbf{H}}_{d',f} \quad t = \arg\max_{m} \mathbf{c}_m, \quad \boldsymbol{\tau}_{n,d'} = t-(M+1).
$$

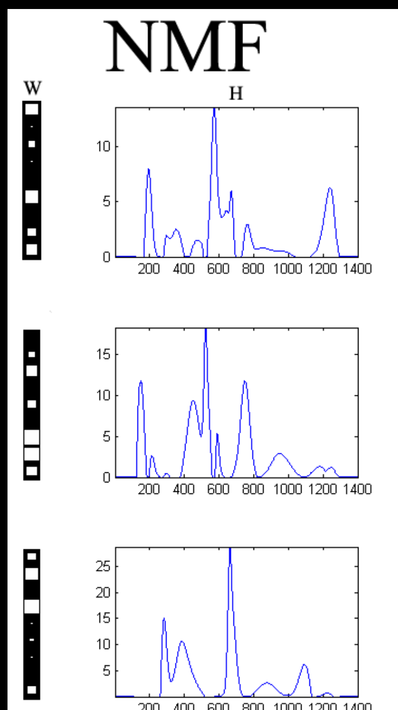
The value of $\mathbf{W}_{n,d'}$ corresponding to this delay is given by

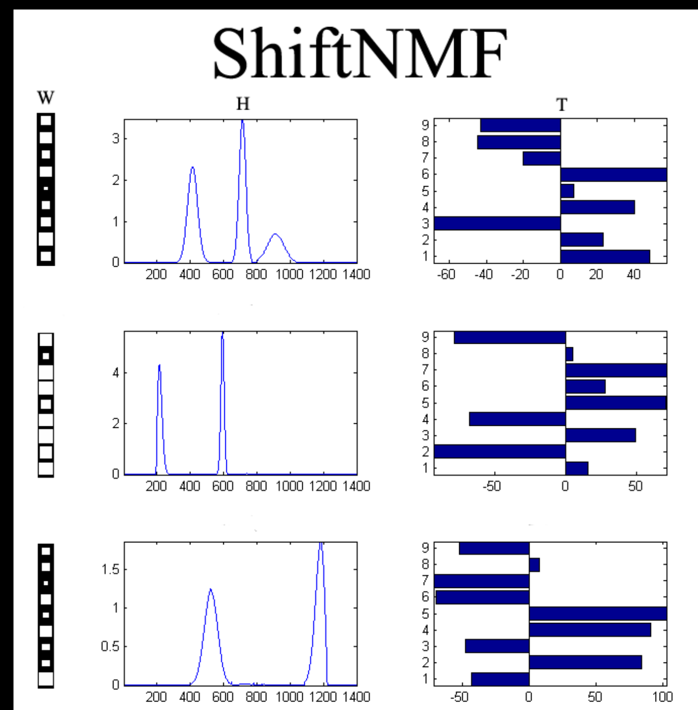
$$
\mathbf{W}_{n,d'} = \frac{\mathbf{c}_t}{\mathbf{H}_{d',:}\mathbf{H}^T_{d',:}}.
$$



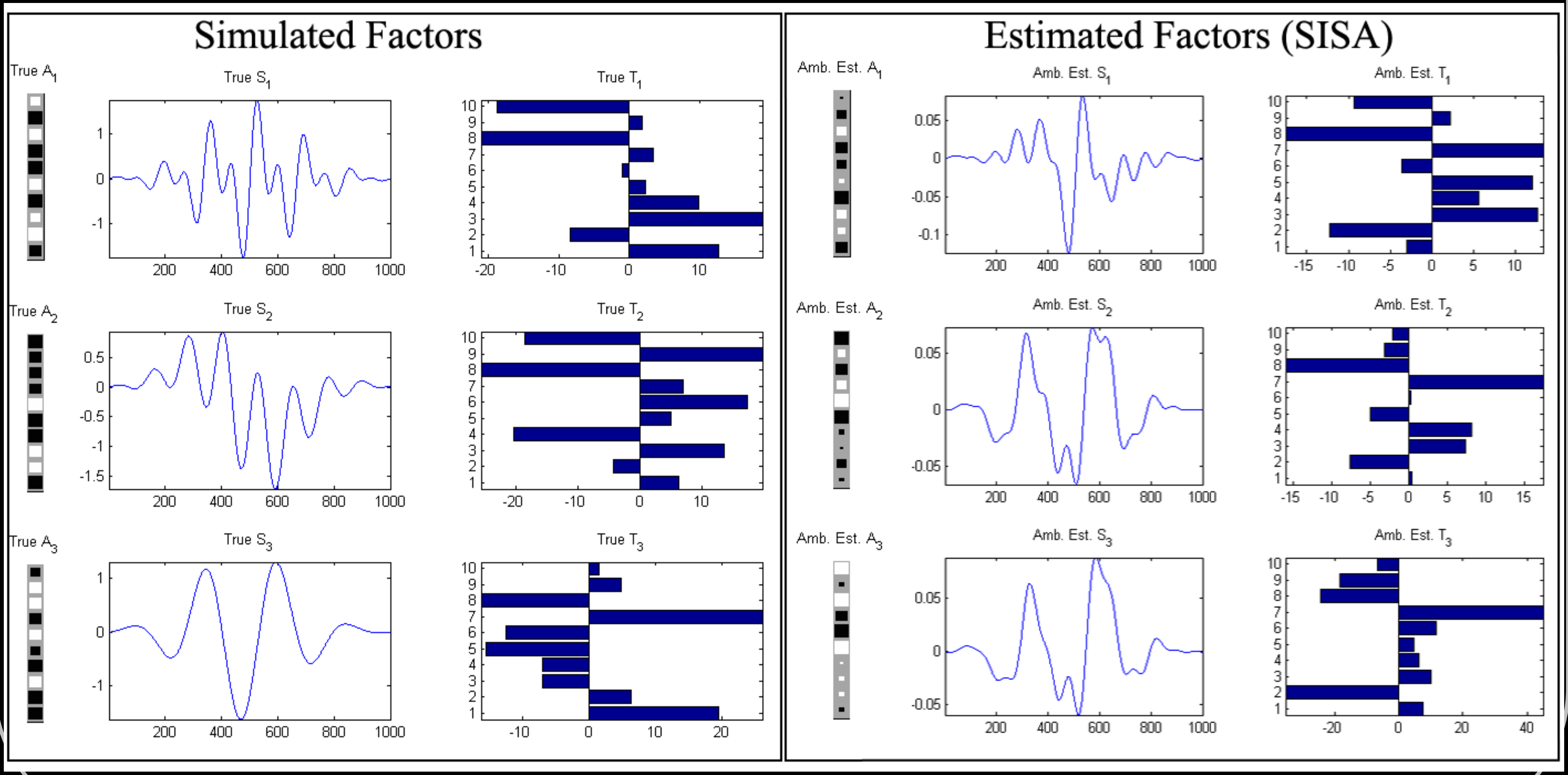

W
H
T


Sensor 1
Sensor 2
Sensor 3
Sensor 4
Sensor 5
Sensor 6
Sensor 7
Sensor 8
Sensor 9


NMF
W
H


ShiftNMF
W
H
T

# Shift Invariant Subspace Analysis



# Shifted Independent Component Analysis

Define, $\widetilde{\mathbf{U}}_f = \widetilde{\mathbf{Q}}^{(f)}\widetilde{\mathbf{H}}_f$, i.e. the sources at frequency $f$ when transformed according to the rotation and shift ambiguity described in the previous section. The ambiguity can be resolved by maximizing the log-likelihood assuming the (non-gaussian) Laplace distribution $p(\widetilde{\mathbf{U}}_f) \propto e^{-|\widetilde{\mathbf{U}}_{d,f}|}$, i.e.

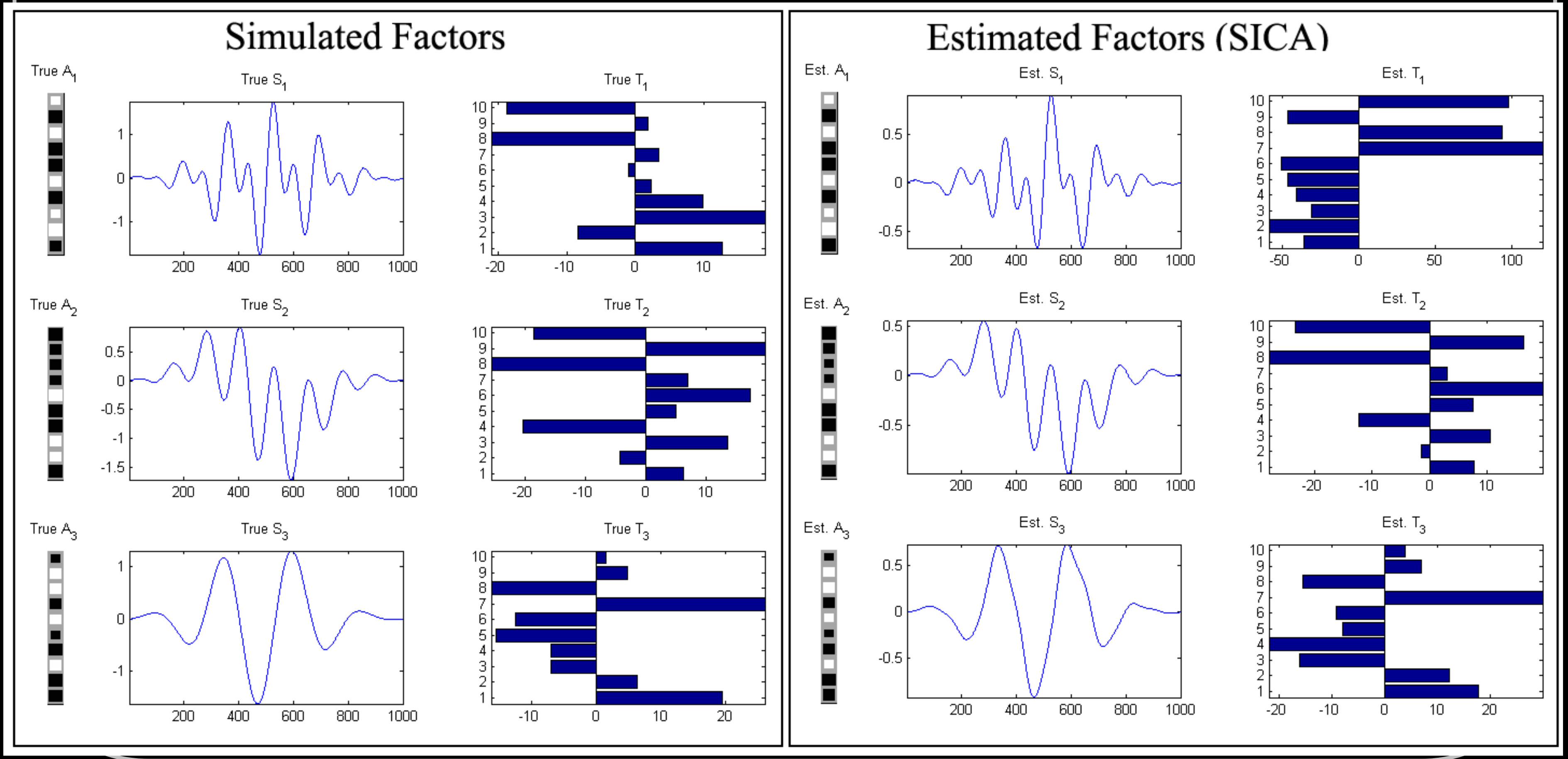
$$p(\widetilde{\mathbf{H}}_f|\mathbf{Q},\widehat{\boldsymbol{\tau}}) = \prod_f p(\widetilde{\mathbf{H}}_f|\mathbf{Q},\widehat{\boldsymbol{\tau}}) = \prod_f |det(\widetilde{\mathbf{Q}}^{(f)})| p(\widetilde{\mathbf{Q}}^{(f)}\widetilde{\mathbf{H}}_f) \tag{1}$$

Such that the log-likelihood as a function of $\mathbf{Q}$ and $\widehat{\boldsymbol{\tau}}$ becomes

$$\mathcal{L}(\mathbf{Q},\widehat{\boldsymbol{\tau}}) = \sum_f \ln|\det(\widetilde{\mathbf{Q}}^{(f)})| - \sum_d |\widetilde{\mathbf{Q}}^{(f)}\widetilde{\mathbf{H}}_f|_d \tag{2}$$

By maximizing $\mathcal{L}(\mathbf{W},\widehat{\boldsymbol{\tau}})$ $\mathbf{W}$ and $\widehat{\boldsymbol{\tau}}$ is estimated and a new unambiguous $\mathbf{H}$ solution found by $\widetilde{\mathbf{H}}_f = \widetilde{\mathbf{Q}}^{(f)}\widetilde{\mathbf{H}}_f$. The corresponding mixing and delays can be estimated alternating between the $\mathbf{W}$ and $\boldsymbol{\tau}$ update. We initialized $\mathbf{W}$ as $\mathbf{W} = \mathbf{W}\mathbf{Q}^{-1}$ and $\boldsymbol{\tau}_{i,d}$ by the cross-correlation procedure.

# Shifted Independent Component Analysis



# Algorithm assumptions

- Sources **H** and measured signal **V** have to be periodic

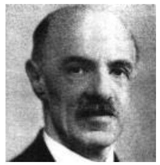
  If signals are not periodic a window function can be employed. However this is not trivial to implement in the $\tau$-update and slows down the algorithm significantly. Zero padding is simple and fast but introduces a bias towards small delays.

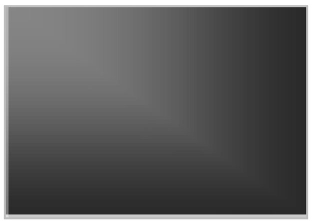
- Noise **E** assumed homoscedatic (normal) iid.

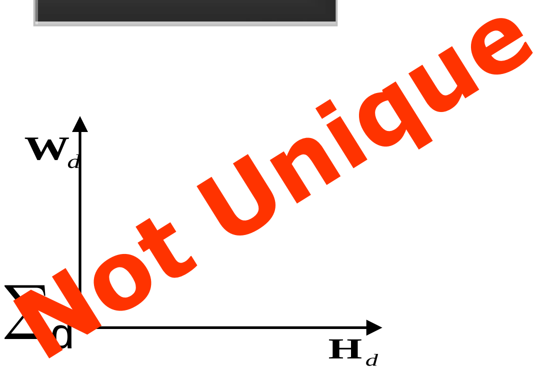
  If non-homoscedatic use weighted least squares. Algorithm works for Least squares due to Parseval's identity. No such identity exists for other types than the least squares objective.

# Extensions to tensors

Factor Analysis

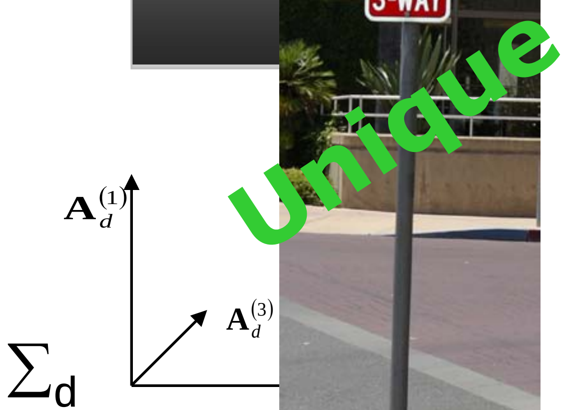







$$\mathbf{V}_{i_1 i_2} \approx \sum_{d=1}^{D} \mathbf{W}_{i_1 d} \mathbf{H}_{i_2 d}$$

PARAFA … mp


$$\mathcal{V}_{i_1 i_2 i_3} \approx \sum_{d=1}^{D} \mathbf{A}_{i_1 d}^{(1)} \ldots$$

# The Candecomp/PARAFAC (CP) model

$$\mathcal{X}_{i,j,k} = \sum_{d}^{D} \mathbf{A}_{i,d}\mathbf{B}_{j,d}\mathbf{C}_{k,d} + \mathcal{E}_{i,j,k}$$

$$\begin{aligned}
\mathbf{X}_{(1)} &= \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T + \mathbf{E}_{(1)} \quad &\Rightarrow \quad \mathbf{A} \leftarrow \mathbf{X}_{(1)}(\mathbf{C} \odot \mathbf{B})^{T^\dagger} \\
\mathbf{X}_{(2)} &= \mathbf{B}(\mathbf{C} \odot \mathbf{A})^T + \mathbf{E}_{(2)} \quad &\Rightarrow \quad \mathbf{B} \leftarrow \mathbf{X}_{(2)}(\mathbf{C} \odot \mathbf{A})^{T^\dagger} \\
\mathbf{X}_{(3)} &= \mathbf{C}(\mathbf{B} \odot \mathbf{A})^T + \mathbf{E}_{(3)} \quad &\Rightarrow \quad \mathbf{C} \leftarrow \mathbf{X}_{(3)}(\mathbf{B} \odot \mathbf{A})^{T^\dagger}
\end{aligned}$$

$$(\mathbf{C} \odot \mathbf{B}) = (\mathbf{C}_1 \otimes \mathbf{B}_1 \ \mathbf{C}_2 \otimes \mathbf{B}_2 \ \ldots \ \mathbf{C}_D \otimes \mathbf{B}_D) \quad X_{(1)} = X^{I \times JK}, \quad X_{(2)} = X^{J \times IK}, \quad X_{(3)} = X^{K \times IJ}$$

The CP model is unique if

$$k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2D + 2$$

where $k_{\mathbf{A}}$ is the k-rank denoting the smallest subset of columns of $\mathbf{A}$ that is guaranteed to be linearly independent. Thus, $k_{\mathbf{A}} \leq rank(\mathbf{A})$.

# Shifted CP model

$$\mathcal{X}_{i,j,k} = \sum_{d}^{D} \mathbf{A}_{i,d} \mathbf{B}_{j-\boldsymbol{\tau}_{k,d},d} \mathbf{C}_{k,d} + \mathcal{E}_{i,j,k}$$

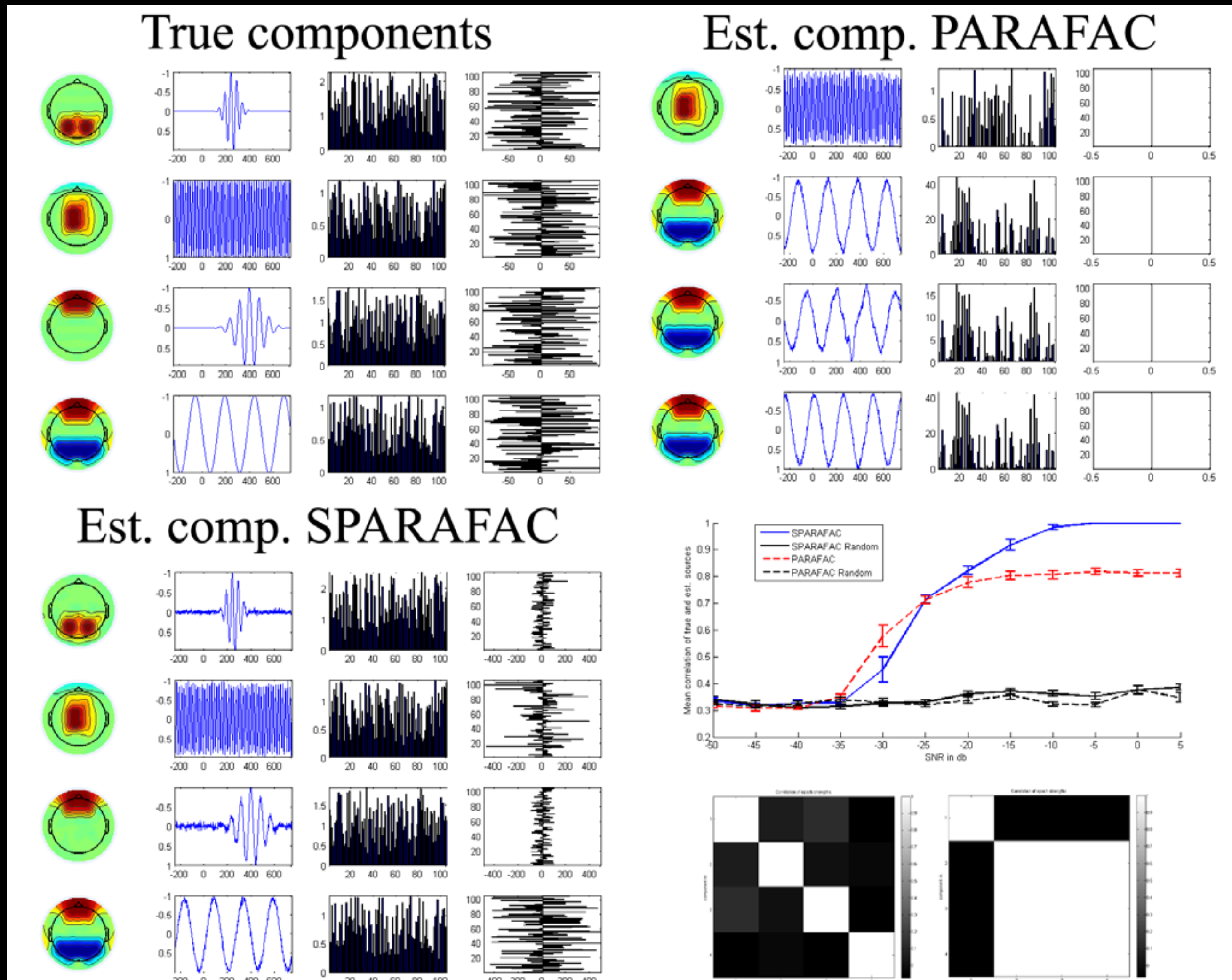
$$\begin{aligned}
\mathbf{X}_{(1)} &= \mathbf{A}\mathbf{Z}^T + \mathbf{E}_{(1)} & \Rightarrow \quad & \mathbf{A} \leftarrow \mathbf{X}_{(1)} \mathbf{Z}^{T^\dagger} \\
\widetilde{\mathbf{X}}_{(2)_{f,:}} &= \widetilde{\mathbf{B}}_{f,:}(\widetilde{\mathbf{C}}^{(f)} \odot \mathbf{A})^T + \widetilde{\mathbf{E}}_{(2)_{f,:}} & \Rightarrow \quad & \widetilde{\mathbf{B}}_{f,:} \leftarrow \widetilde{\mathbf{X}}_{(2)_{f,:}} (\widetilde{\mathbf{C}}^{(f)} \odot \mathbf{A})^{T^\dagger} \\
\mathbf{X}_{(3)_{k,:}} &= \mathbf{C}_{k,:}(\mathbf{B}^{(k)} \odot \mathbf{A})^T + \mathbf{E}_{(3)_{k,:}} & \Rightarrow \quad & \mathbf{C}_{k,:} \leftarrow \mathbf{X}_{(3)_{k,:}} (\mathbf{B}^{(k)} \odot \mathbf{A})^{T^\dagger}
\end{aligned}$$

$$\mathbf{r}_j^{(k,d')} = \sum_i \mathcal{R}_{i,j,k}^{d'} \mathbf{A}_{i,d'} \quad \widetilde{\mathbf{c}}_{k,d'}(f) = \widetilde{\mathbf{r}}_f^{(k,d')^*} \widetilde{\mathbf{B}}_{f,d'}.$$

$$t_{k,d'} = \arg\max_t |\mathbf{c}_{k,d'}(t)| \quad \boldsymbol{\tau}_{k,d'} = t_{k,d'} - (J+1).$$

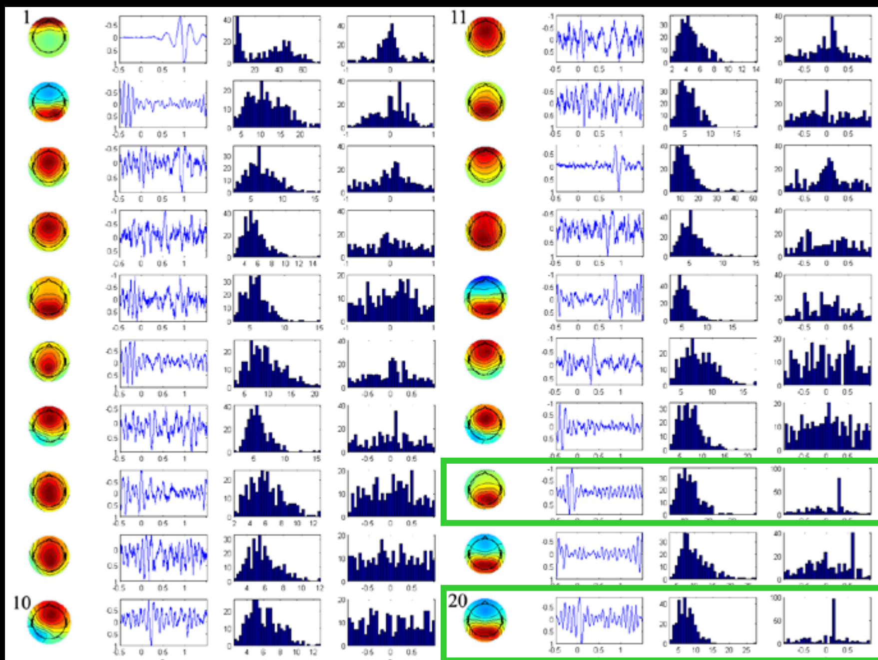
$$\mathbf{C}_{k,d'} = \frac{\mathbf{c}_{k,d'}(t_{k,d'})}{\mathbf{B}_{d'}^T \mathbf{B}_{d'}}.$$

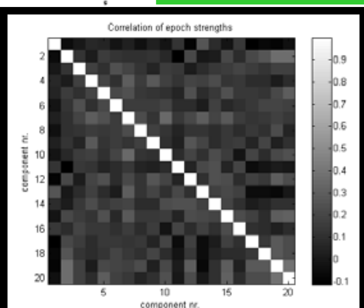
# Synthetic EEG data



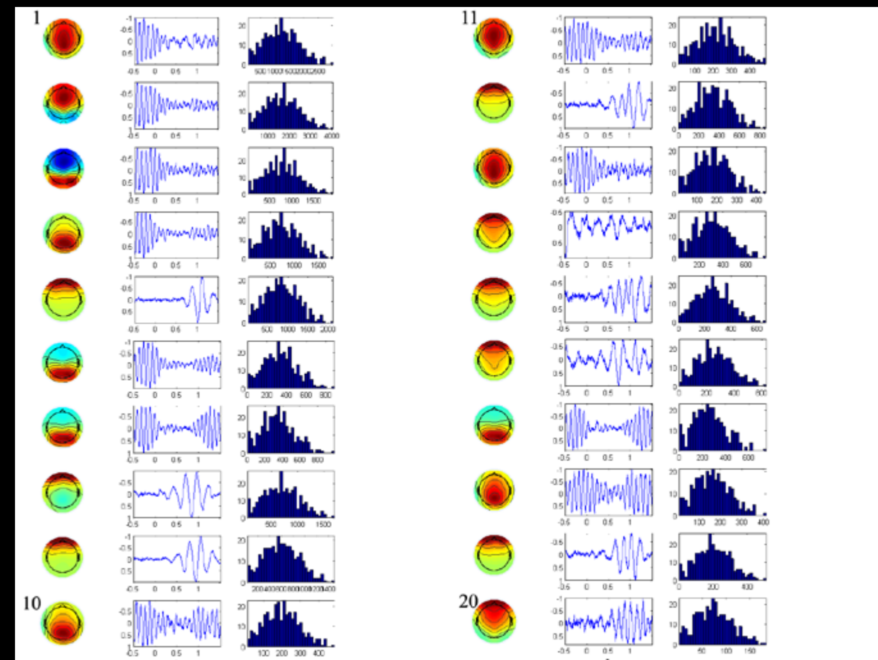
# EEG data from Visual Paradigm

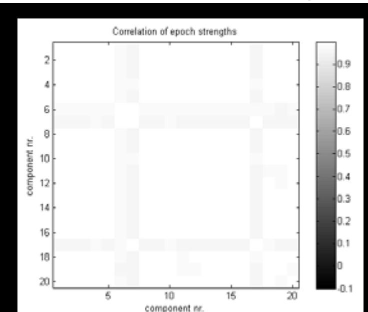
## Shifted CP





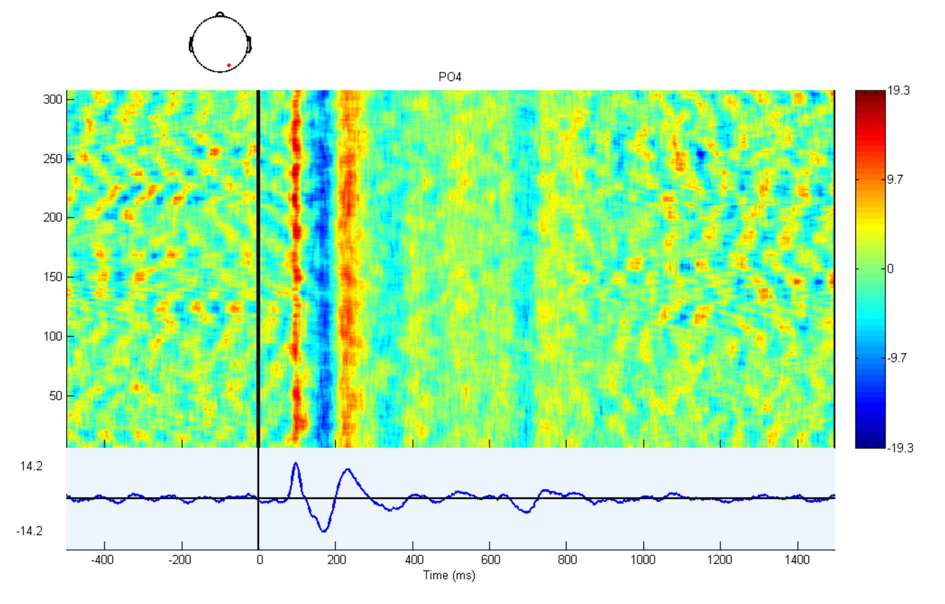

## CP

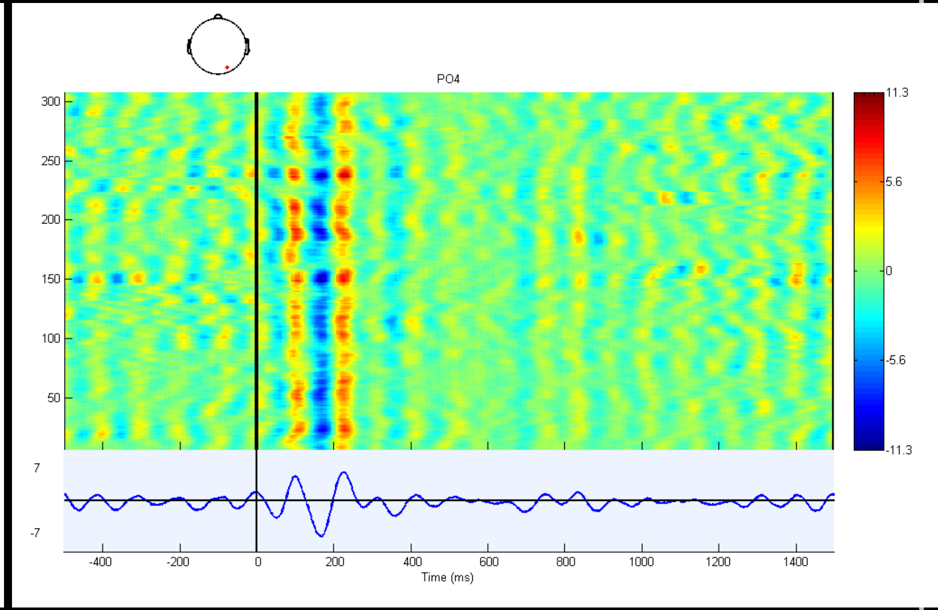


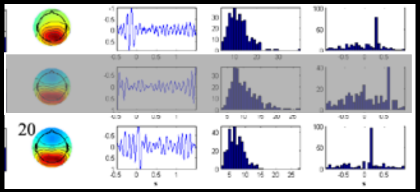


# True Evoked Potential (EP)

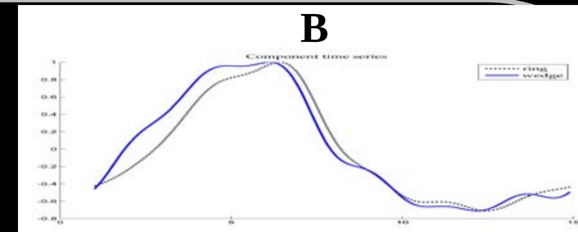
# Reconstructed EP component 18 and 20



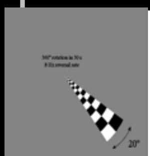



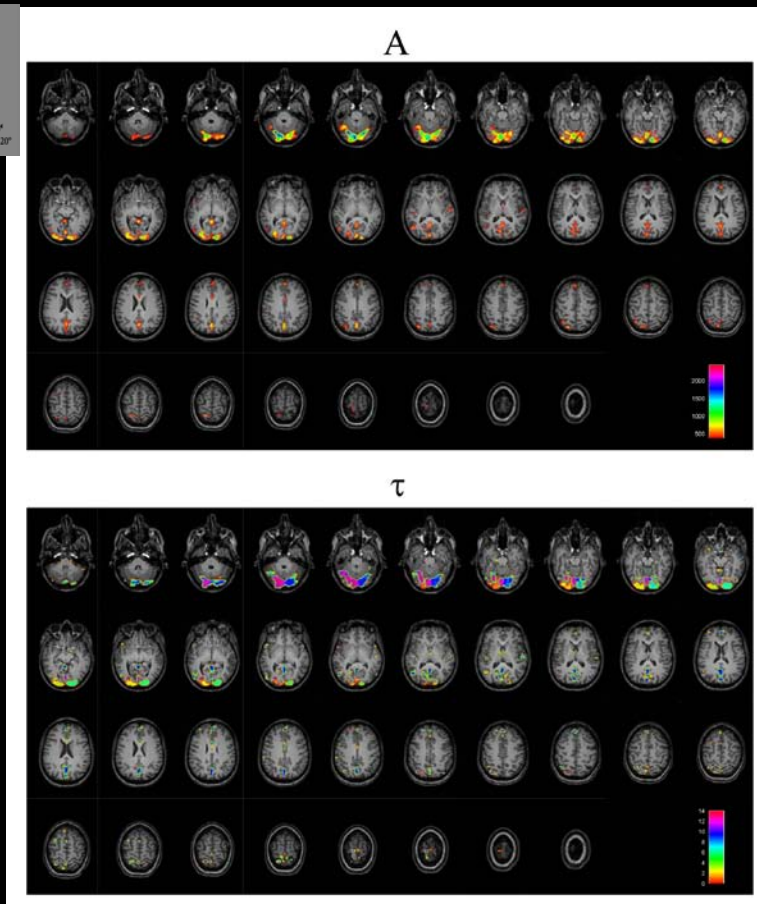



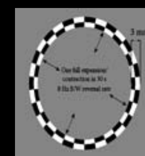
# fMRI data visual paradigm

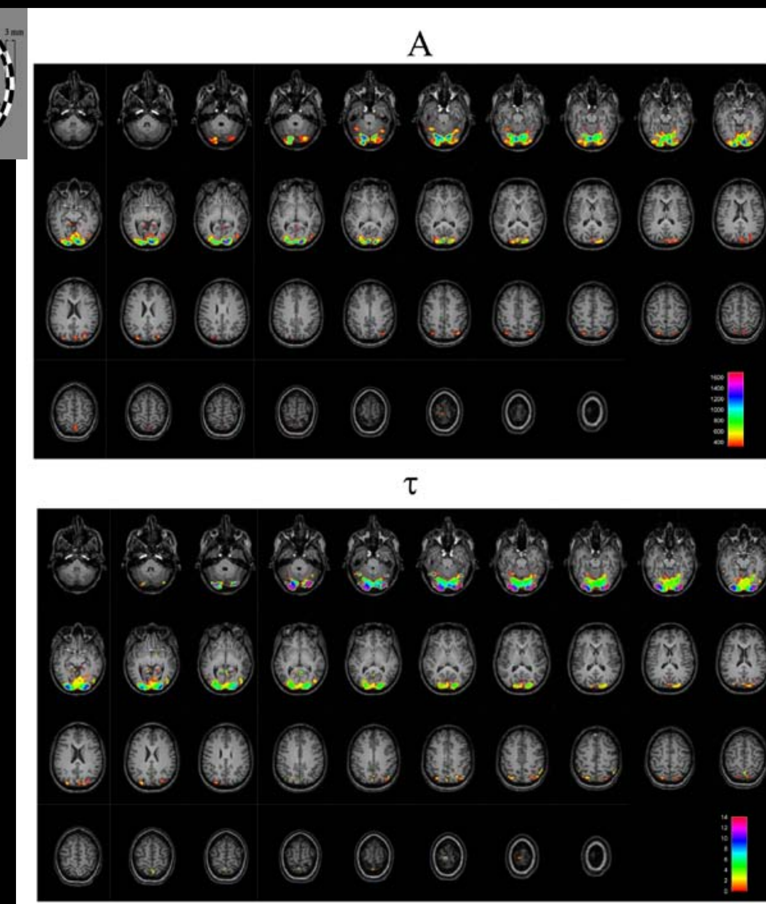



$$\mathcal{X}_{i,j,k} = \sum_{d}^{D} \mathbf{A}_{i,d}\mathbf{B}_{j-\boldsymbol{\tau}_{k,d},d}\mathbf{C}_{k,d} + \mathcal{E}_{i,j,k}$$








(Kristoffer Hougaard Madsen)

# Shift invariant sparse coding

## (Solving the shift problem using sparse coding)

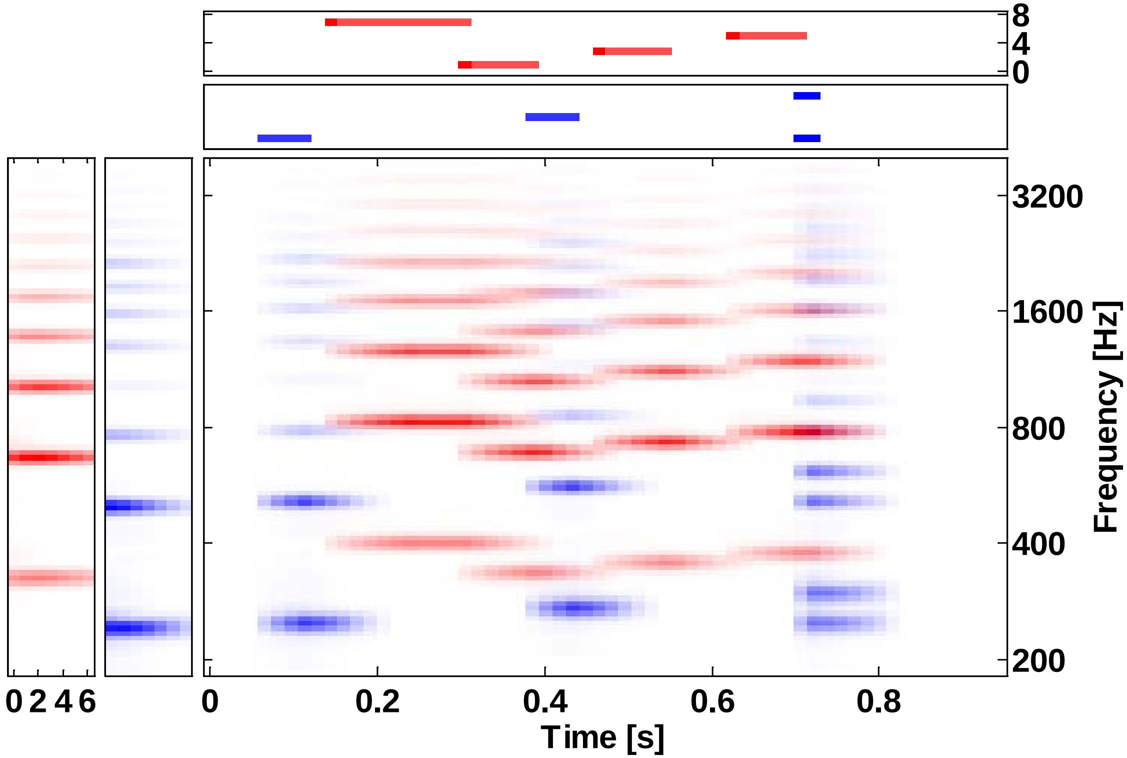
$$X(x,y) \approx \sum_{d} \sum_{u,v} \alpha_d(u,v)\Phi_d(x-u,y-v)$$

**Non-negative matrix 2D de-convolution introduced in:**

**Schmidt and Mørup, 2006d**

**Extended to Sparse coding**

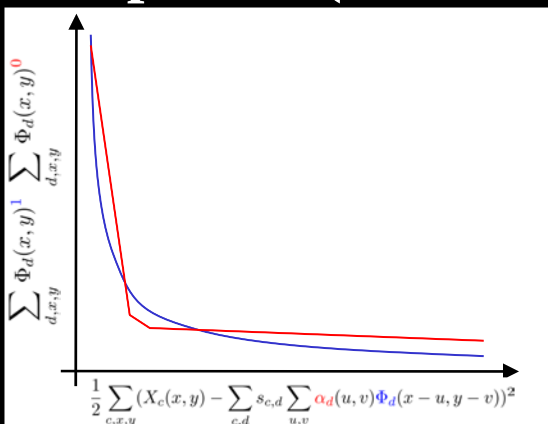
**Mørup et al. 2007b**



Objective function:

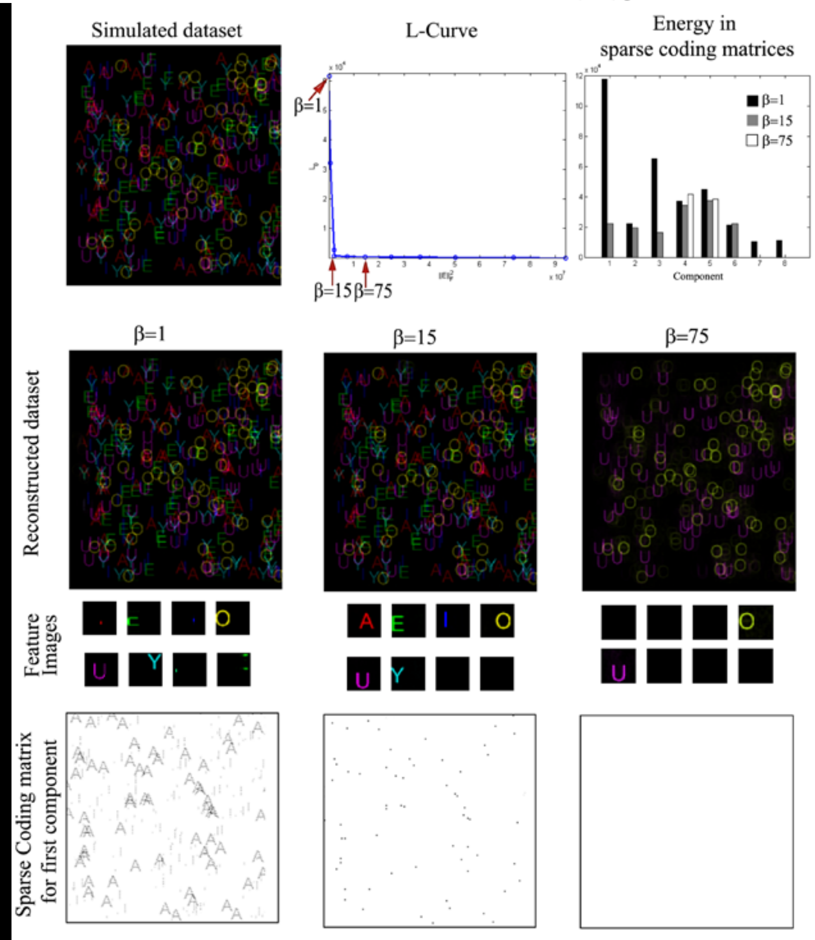
$$\frac{1}{2}\underbrace{\sum_{c,x,y}\Big(X_c(x,y)-\sum_{c,d}s_{c,d}\sum_{u,v}{\color{red}\alpha_d}(u,v){\color{blue}\Phi_d}(x-u,y-v)\Big)^2}_{\textbf{Reconstruction}}+\beta\underbrace{\sum_{d,x,y}\Phi_d(x,y)}_{\mathbf{L_1\ norm}}$$

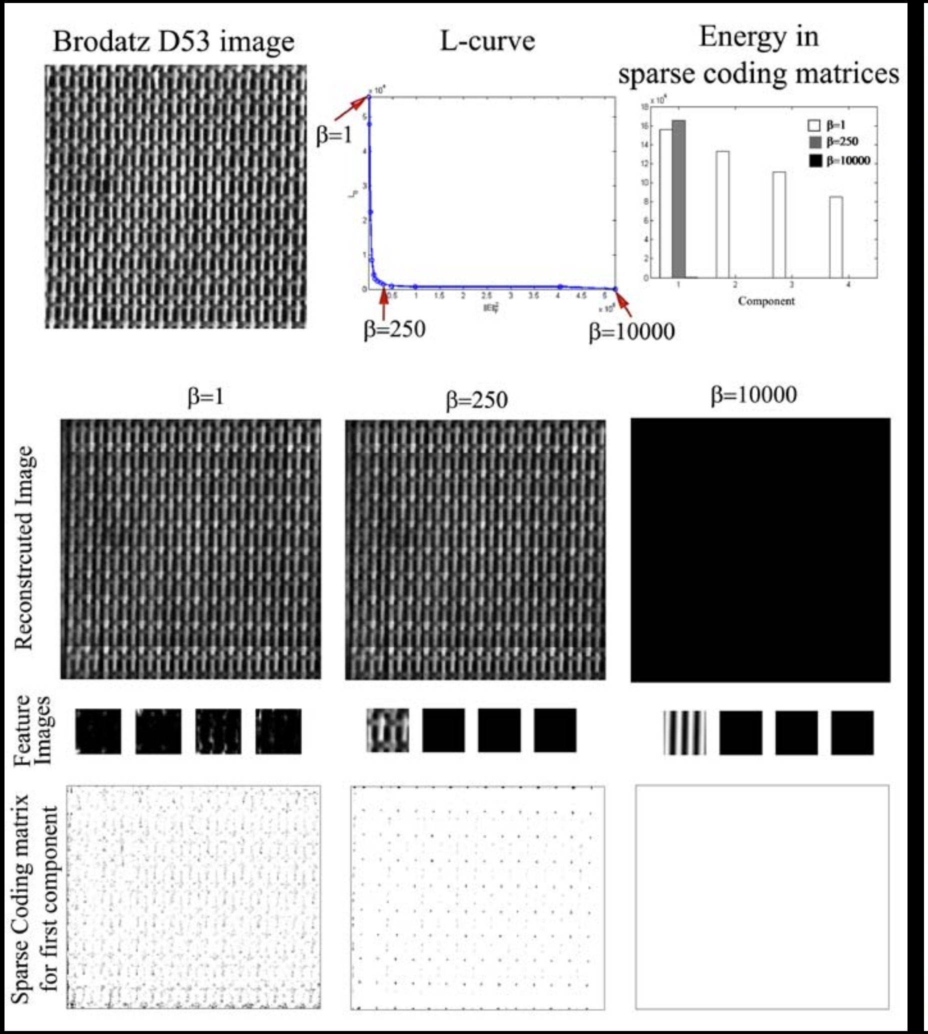
**Solved by multiplicative updates (i.e., X, s, α, Φ≥0)**

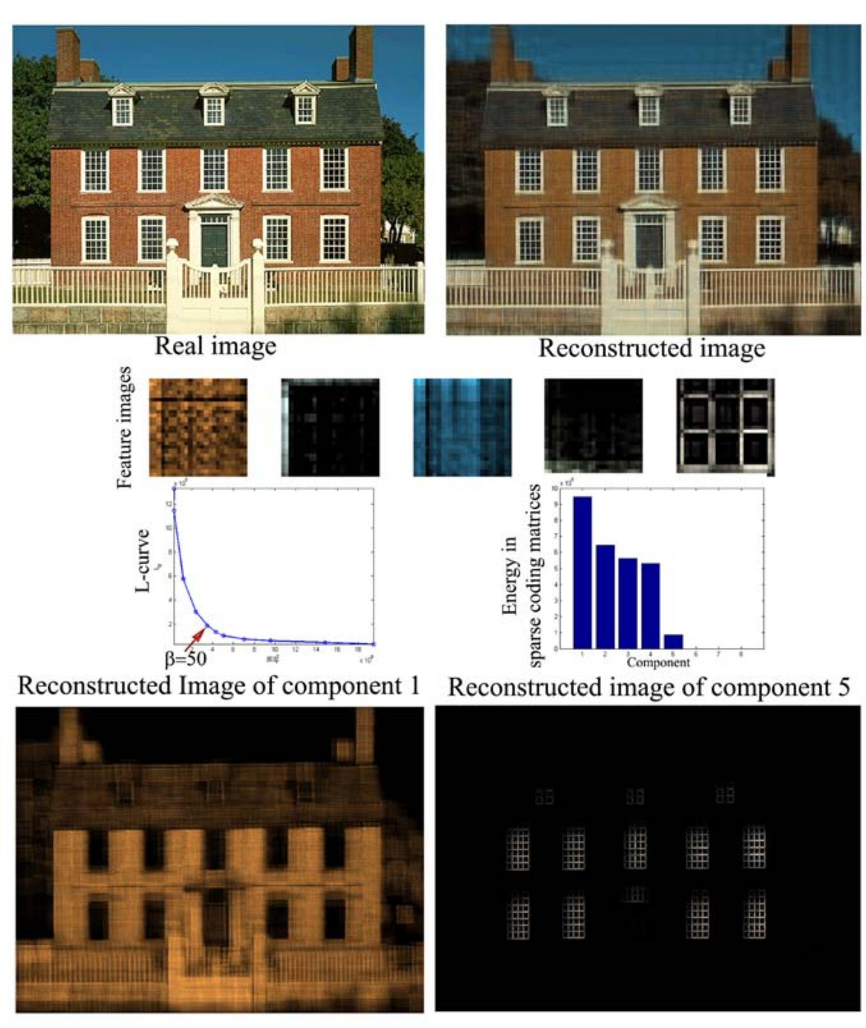


$$I_c(u,v)=s_{c,d}{\color{red}\alpha_d}(u,v)\rightarrow$$

$${\color{blue}\Phi_d}(x-u,y-v)\rightarrow$$



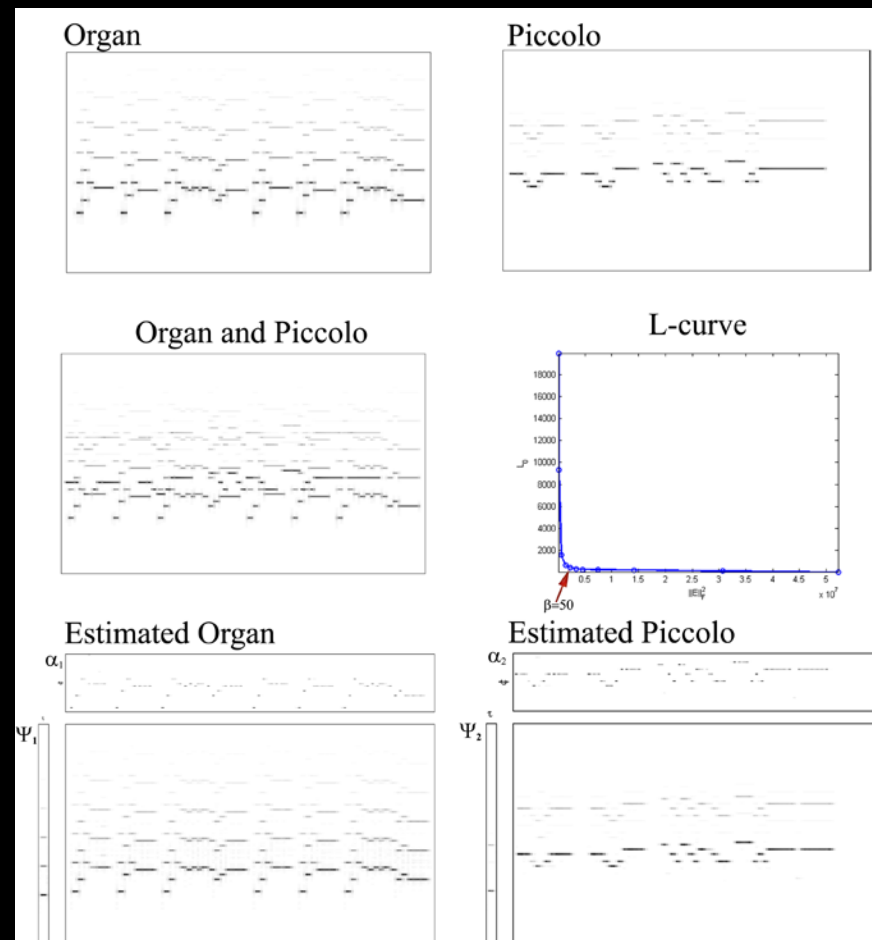

Brodatz D53 image
L-curve
Energy in
sparse coding matrices
β=1
β=250
β=10000
Component
β=1
β=250
β=10000
Reconstructed Image
Feature
Images
Sparse Coding matrix
for first component


Real image
Reconstructed image
Feature images
L-curve
β=50
Energy in
sparse coding matrices
Component
Reconstructed Image of component 1
Reconstructed image of component 5

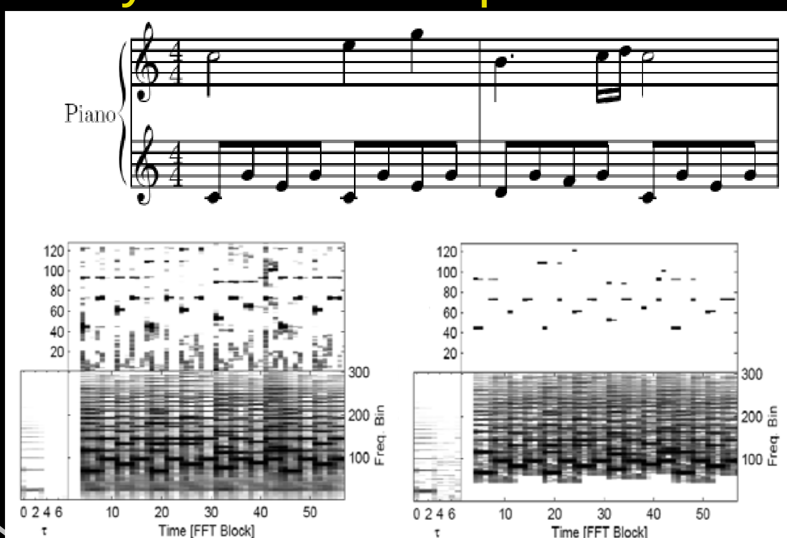
# Analysis of mono signal of mixed Organ and Piccolo

(Music data taken from Y.-G. Zhang, 2005)

$$X(x,y) \approx \sum_{d,u,v} \alpha_d(u,v)\Psi_d(x-u,y-v)$$




## Analysis of mono piano music



## Conclusion:

Modelling data using shift invariance seems highly relevant for a wide range of data types.

Two approaches presented: 1) estimation of specific shift in the frequency domain 2) Estimation of shifts using sparse coding.

Degeneracies encountered in the Candecomp/PARAFAC (CP) model appear to vanish when allowing for shifts. Thus degeneracy in CP often a result of sources of the data being shifted.


- Mørup, M., Hansen, L. K., Arnfred, S. M., Lim, L., Madsen, K. H., *Shift Invariant Multilinear Decomposition of Neuroscience Data*, 2007
- M. Mørup, M. N. Schmidt, L. K. Hansen, *Shift Invariant Sparse Coding of Image and Music Data*, submitted, JMLR, 2007
- M. Mørup, K. H. Madsen, L. K. Hansen, *Shifted Non-negative Matrix Factorization*, MLSP 2007, 2007
- M. Mørup, K. H. Madsen, L. K. Hansen, *Shifted Independent Component Analysis*, ICA2007, 2007

# References


Carroll, J. D. and Chang, J. J. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition, Psychometrika 35 1970 283—319

Donoho, D. and Stodden, V. When does non-negative matrix factorization give a correct decomposition into parts? NIPS2003

Donoho D. For most large underdetermined systems of linear equations the minimal $l1$-norm solution is also the sparsest solution. Communications on Pure and Applied Mathematics, 59(6):797829, 2006.

Eggert, J. and Korner, E. Sparse coding and NMF. In Neural Networks volume 4, pages 2529-2533, 2004

FitzGerald, D., Cranitch, M., and Coyle, E., "Shifted 2D Non-negative Tensor Factorisation", *Proceedings of the Irish Signals and Systems Conferenc*e, Dublin, June 2006.

Kruskal, J.B. Three-way analysis: rank and uniqueness of trilinear decompostions, with application to arithmetic complexity and statistics. Linear Algebra Appl., 18: 95-138, 1977

G.H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics, 21(2):215223, 1979.

L. K. Hansen, K. H. Madsen, and T. Lehn-Schiøler. Adaptive regularization of noisy linear inverse problems. In Eusipco, 2006.

Harshman, R. A. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis},UCLA Working Papers in Phonetics 16 1970 1—84

Harshman, Richard A.Harshman and Hong, Sungjin Lundy, Margaret E. Shifted factor analysis—Part I: Models and properties J. Chemometrics (17) pages 379–388, 2003

Harshman, R.A., S. Hong, and M.E. Lundy, Shifted factor analysis part II: Algorithms, Journal of Chemometrics, vol. 17, pp. 379-388, 2003.

Hoyer, P.O. Non-negative matrix factorization with sparseness constraints. Journal of Machine Learning Research, 2004.

Lathauwer, Lieven De and Moor, Bart De and Vandewalle, Joos MULTILINEAR SINGULAR VALUE DECOMPOSITION.SIAM J. MATRIX ANAL. APPL.2000 (21)1253–1278

Lee, D.D. and Seung, H.S. Algorithms for non-negative matrix factorization. In NIPS, pages 556-462, 2000

Lee, D.D and Seung, H.S. Learning the parts of objects by non-negative matrix factorization, NATURE 1999

Lim, Lek-Heng - http://www.etis.ensea.fr/~wtda/Articles/wtda-nnparafac-slides.pdf

Lim, L.-H. and Golub, G.H., "Nonnegative decomposition and approximation of nonnegative matrices and tensors," SCCM Technical Report, 06-01, forthcoming, 2006.

Murakami, Takashi and Kroonenberg, Pieter M. Three-Mode Models and Individual Differences in Semantic Differential Data, Multivariate Behavioral Research(38) no. 2 pages 247-283, 2003

Mørup, M. and Hansen, L.K.and Arnfred, S.M.Decomposing the time-frequency representation of EEG using nonnegative matrix and multi-way factorization Technical report, Institute for Mathematical Modeling, Technical University of Denmark, 2006b

Mørup, M., Hansen, L. K., Arnfred, S. M., ERPWAVELAB A toolbox for multi-channel analysis of time-frequency transformed event related potentials, *Journal of Neuroscience Methods*, vol. 161, pp. 361-368, 2007a

Mørup, M., Hansen, L. K., Parnes, Josef, Hermann, C, Arnfred, S. M., Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG Neuroimage NeuroImage 29 938 – 947, 2006a

Mørup, M., Schmidt, M. N., Hansen, L. K., *Shift Invariant Sparse Coding of Image and Music Data*, submitted, JMLR, 2007b

Mørup, M., Hansen, L. K., Arnfred, S. M., *Algorithms for Sparse Non-negative TUCKER*, Submitted Neural Computation, 2006c

Mørup, M, Madsen, K.H., Hansen, L.K. Shifted Non-negative Matrix Factorization, to appear MLSP 2007c

Nørgaard, L and Ridder, C.Rank annihilation factor analysis applied to flow injection analysis with photodiode-array detection Chemometrics and Intelligent Laboratory Systems 1994 (23) 107-114

Schmidt, M.N. and Mørup, M. Non-negative matrix factor 2D deconvolution for blind single channel source separation. In ICA2006, pages 700-707, 2006d

Smaragdis, P. Non-negative Matrix Factor deconvolution; Extraction of multiple sound sources from monophonic inputs. International Symposium on independent Component Analysis and Blind Source Separation (ICA)

Smilde, Age K. Smilde and Tauller, Roma and Saurina, Javier and Bro, Rasmus, Calibration methods for complex second-order data Analytica Chimica Acta 1999 237-251

Sun, Jian-Tao and Zeng, Hua-Jun and Liu, Huanand Lu Yuchang and Chen Zheng CubeSVD: a novel approach to personalized Web search WWW '05: Proceedings of the 14th international conference on World Wide Web pages 382—390, 2005

Tamara G. Kolda Multilinear operators for higher-order decompositions technical report Sandia national laboratory 2006 SAND2006-2081.

Tucker, L. R. Some mathematical notes on three-mode factor analysis Psychometrika 31 1966 279—311

Welling, M. and Weber, M. Positive tensor factorization. Pattern Recogn. Lett. 2001

Vasilescu , M. A. O. and Terzopoulos , Demetri Multilinear Analysis of Image Ensembles: TensorFaces, ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part I, 2002

Y.-G. Zhang and C.-S. Zhang. Separtation of music signals by harmonic structure modeling. Proceedings of Neural Information Processing Systems (NIPS), pages 184191, 2005.

# References cont.


Anthony J. Bell and Terrence J. Sejnowski, An information maximization approach to blind source separation and blind deconvolution, Neural Computation, vol. 7, pp. 11291159, 1995.

B. Bödvarsson, L. K. Hansen, C. Svarer, and G. M. Knudsen, NMF on positron emission tomography, in International conference on acoustics, speech and signal processing 2007, ICASSP., 2007.

Shuyan Du, Paul Sajda, Radka Stoyanova, and Truman. R. Brown, Recovery of metabolomic spectral sources using non-negative matrix factorization, Proceedings of the 2005 IEEE Engineering in Medicine and Biology, pp. 4731 4734, 2005.

J. Eggert and E. Korner, Sparse coding and nmf, in Neural Networks, 2004, vol. 4, pp. 25292533.

Bruno Emile and Pierre Comon, Estimation of time delays between unknown colored signals, Signal Processing, vol. 68, no. 1, pp. 93100, 1998.

Cyril Gobinet, Eric Perrin, and Régis Huez, Application of non-negative matrix factorization to fluorescence spectroscopy, EUSIPCO, pp. 10951098, 2004.

R.A. Harshman, S. Hong, and M.E. Lundy, Shifted factor analysis.part i: Models and properties, Jour- nal of Chemometrics, vol. 17, pp. 363378, 2003.

R.A. Harshman, S. Hong, and M.E. Lundy, Shifted factor analysis.part ii: Algorithms, Journal of Chemometrics, vol. 17, pp. 379-388, 2003.

Kyeong Min Kim, H Watabe, M Shidahara, Ji Young Ahn, Seungjin Choi, N Kudomi, K Hayashida, Y Miyake, and H Iida, Noninvasive estimation of cerebral blood ow using image-derived carotid input function in h15 2 o dynamic pet, 2001, vol. 3, pp. 1282-1285.

D.D. Lee and H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature,vol. 401, no. 6755, pp. 788-91, 1999.

Daniel D Lee and H. Sebastian Seung, Algorithms for non-negative matrix factorization, in NIPS, 2000, pp. 556562.

J. S. Lee Lee, D. D. Lee, S. Choi, K. S. Park, and D. S. Lee, Non-negative matrix factorization of dynamic images in nuclear medicine, IEEE Nuclear Science Symposium and Medical Imaging Conference, pp. 20272030, 2001.

Morten Mørup and Kristoer H. Madsen, Shift nmf algorithm, http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/5232/zip/imm5232.zip.

Morten Mørup, Kristoer H. Madsen, and Lars K. Hansen, Shifted independent component analysis, to appear, ICA2007, 2007.

P Paatero and U Tapper, Positive matrix factorization:A non-negative factor model with optimal utilization of error estimates of data values, Environmetrics, vol. 5,no. 2, pp. 111126, 1994.

V. Paul Pauca, J. Piper, and Robert J. Plemmons, Nonnegative matrix factorization for spectral data analysis, Linear Algebra and its Applications, vol. 416 (2006), pp. 2947, 2006.

P.O. Hoyer, Non-negative matrix factorization with sparseness constraints, Journal of Machine Learning Research, 2004.

Kaare B. Petersen and Michael S. Pedersen, The matrix cookbook, www.matrixcookbook.com, 2007.

Paris Smaragdis, Non-negative matrix factor deconvolution; extraction of multiple sound sourses from monophonic inputs, International Symposium on In- dependent Component Analysis and Blind Source Separation (ICA), vol. 3195, pp. 494, sep 2004.

K Torkkola, Blind separation of delayed sources based on informationmaximization, Acoustics, Speech, and Signal Processing. ICASSP-96, vol. 6, pp. 3509-3512, 1996.

Arie Yeredor, Time-delay estimation in mixtures, ICASSP, vol. 5, pp. 237240, 2003.

Arie Yeredor, Blind source separation in the presence of doppler frequency shifts, ICASSP, vol. 5, pp. 277 280, 2005.