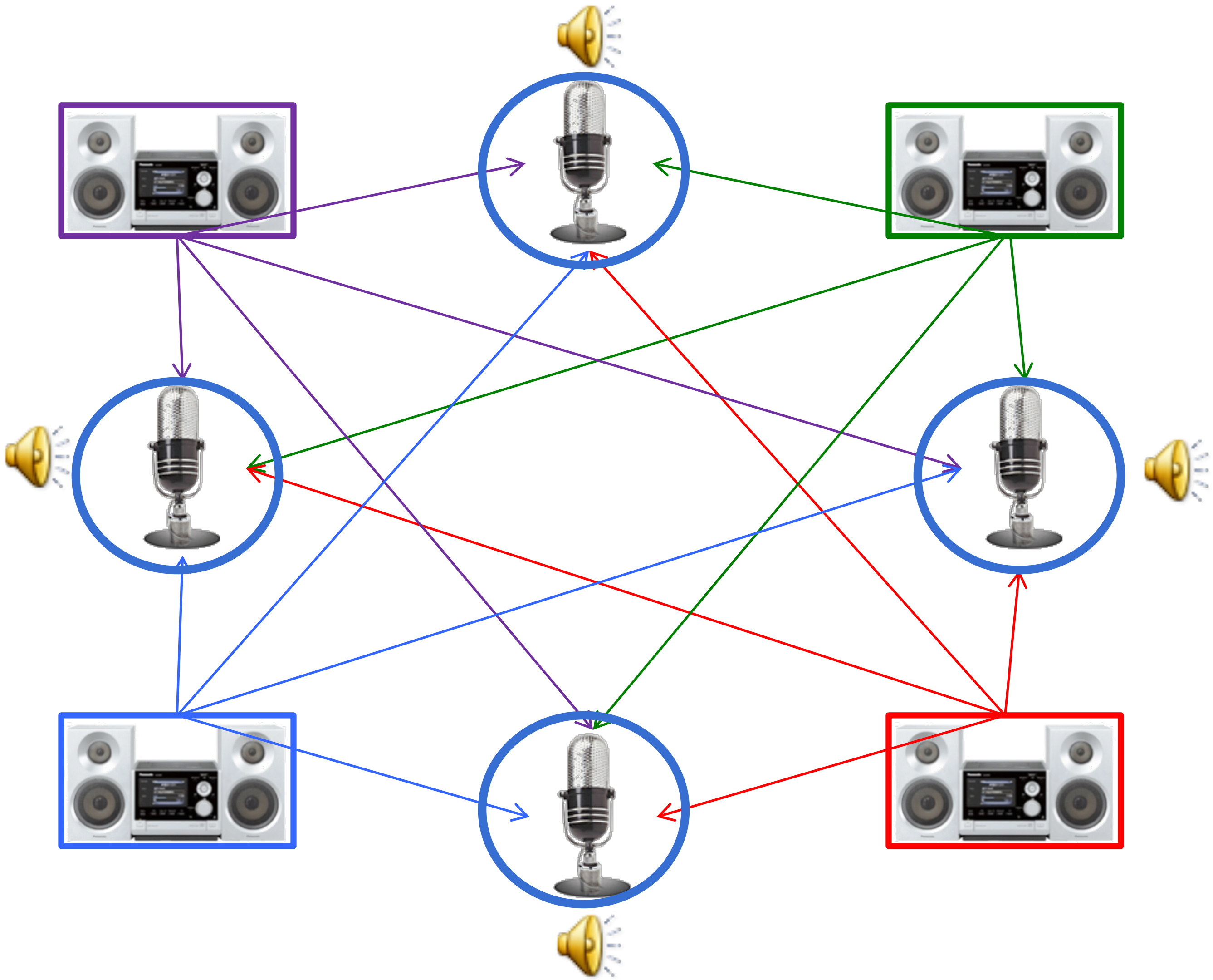
Week 14 Dimensionality Reduction

L14-03 Independent Component Analysis: Concept

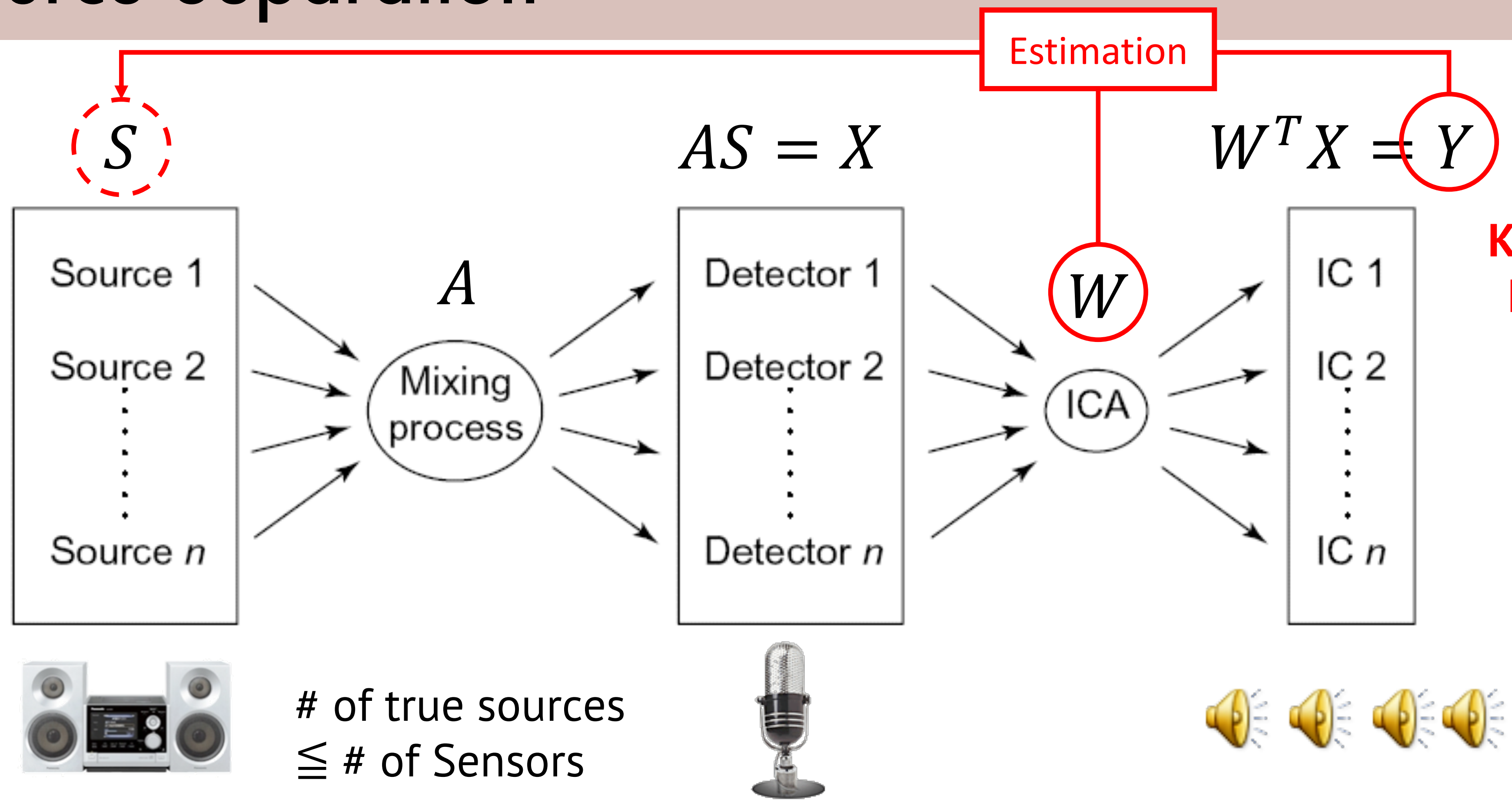
Dong Hee Lee
Master's Student

(Blind) Source Separation



(NeWBI 4 fMRI)

(Blind) Source Separation



Key Question:
How does Y relate to S?

True sources
(Latent constructs)

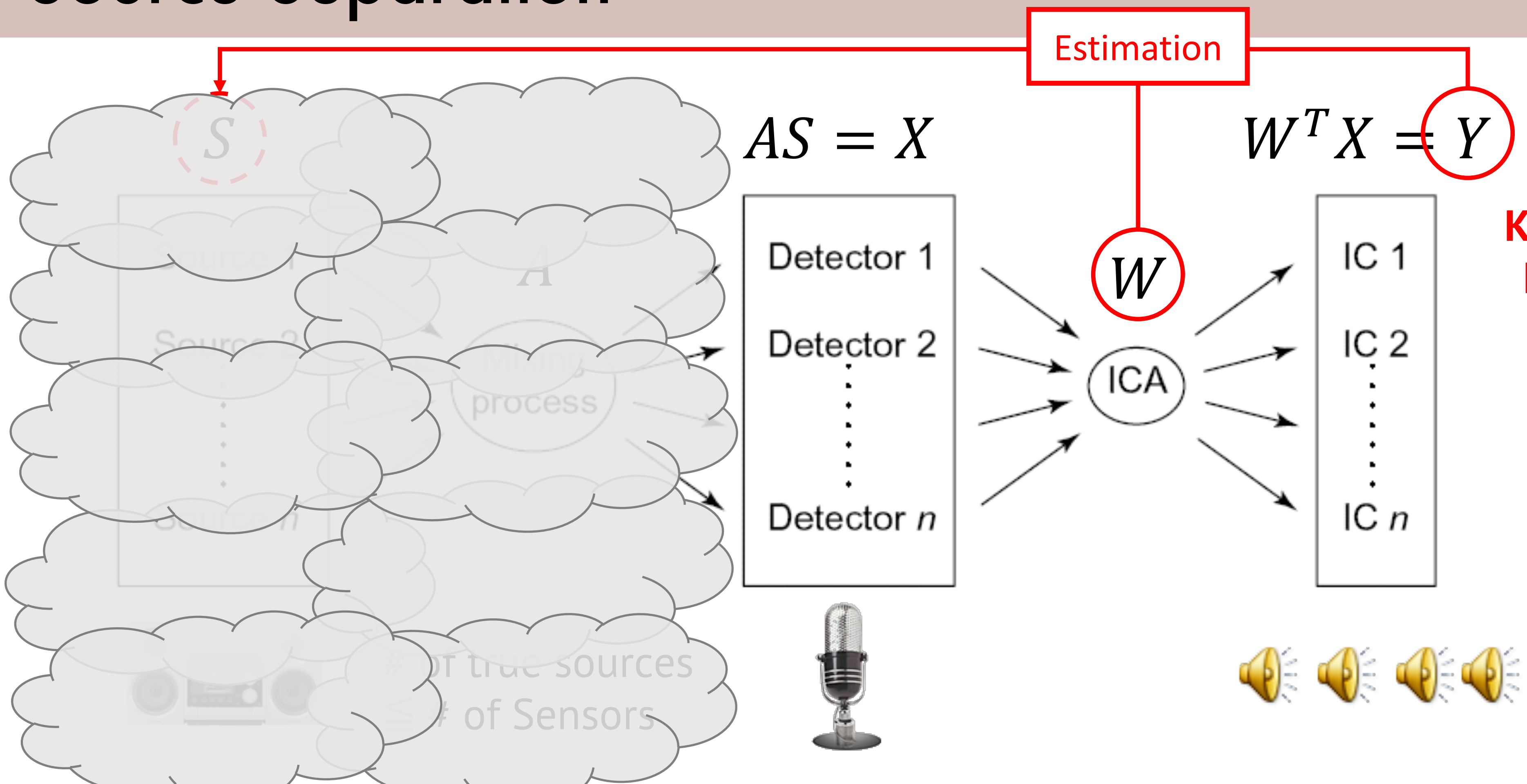
Sensors / Observations
(Manifest variables)

Mixed and overlapped

Sources
(Components)

(NeWBI 4 fMRI)

(Blind) Source Separation



Key Question:
How does Y relate to S?

True sources
(Latent constructs)

Sensors / Observations
(Manifest variables)

Mixed and overlapped

Sources
(Components)

(NeWBI 4 fMRI)

(Blind) Source Separation

Key Question: How does Y relate to S?

$$AS = X$$

$$W^T X = Y$$

If $Y = S$ then

$$\left.\begin{aligned} AY &= X \\ W^T X &= Y \end{aligned}\right\} \quad \boxed{(W^T)^{-1} = A}$$

"Weights matrix"
"Unmixing matrix"

"Forward model"
"Activation pattern"

(Mike X Cohen)

ICA Assumption

- **Assumption #1**: (Statistical) Independent true sources
- **Assumption #2**: Truly linear mixing
- **Assumption #3**: Non-Gaussian distribution of true sources

ICA Assumption

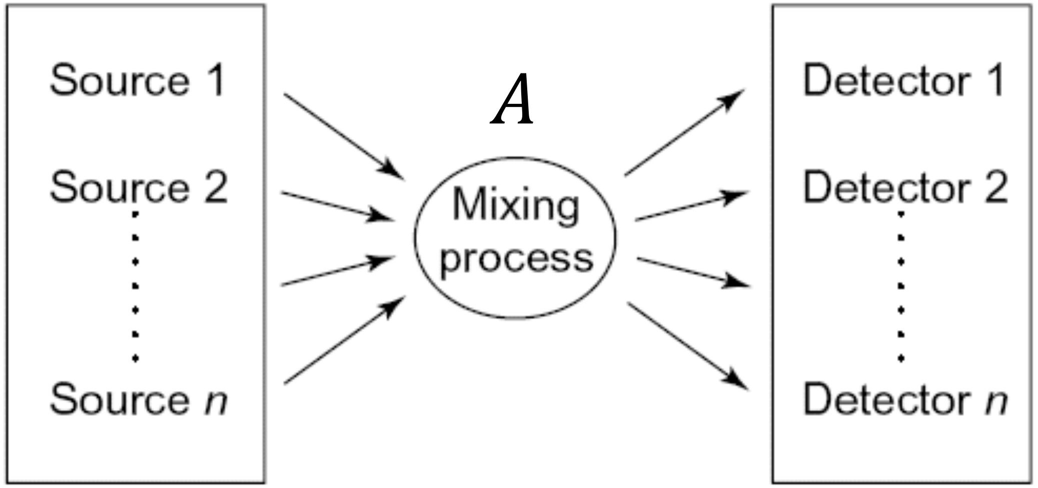
- **Assumption #1**: (Statistical) Independent true sources
  - Y_1, Y_2 are independent $\Leftrightarrow p(y_1, y_2) = p(y_1)p(y_2)$
  - Y_1, Y_2 are independent $\underset{\nLeftarrow}{\Rightarrow}$ Y_1, Y_2 are uncorrelated

- **Assumption #2**: Truly linear mixing
- **Assumption #3**: Non-Gaussian distribution of true sources

ICA Assumption

- **Assumption #1**: (Statistical) Independent true sources
- **Assumption #2**: Truly linear mixing

$$AS = X$$

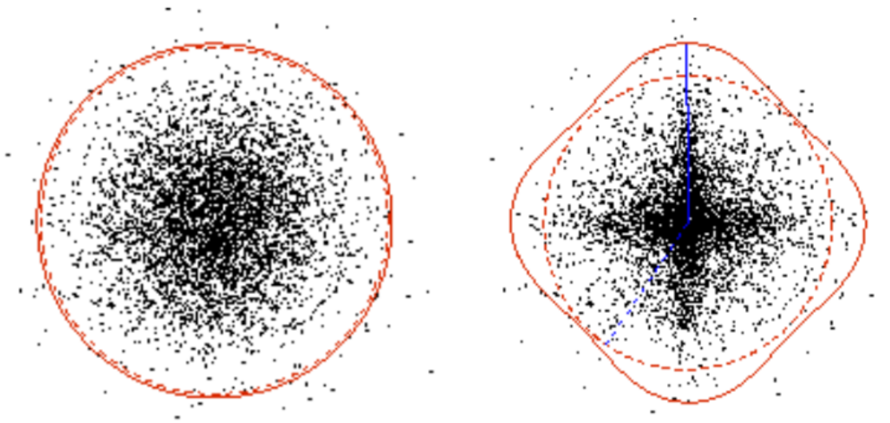


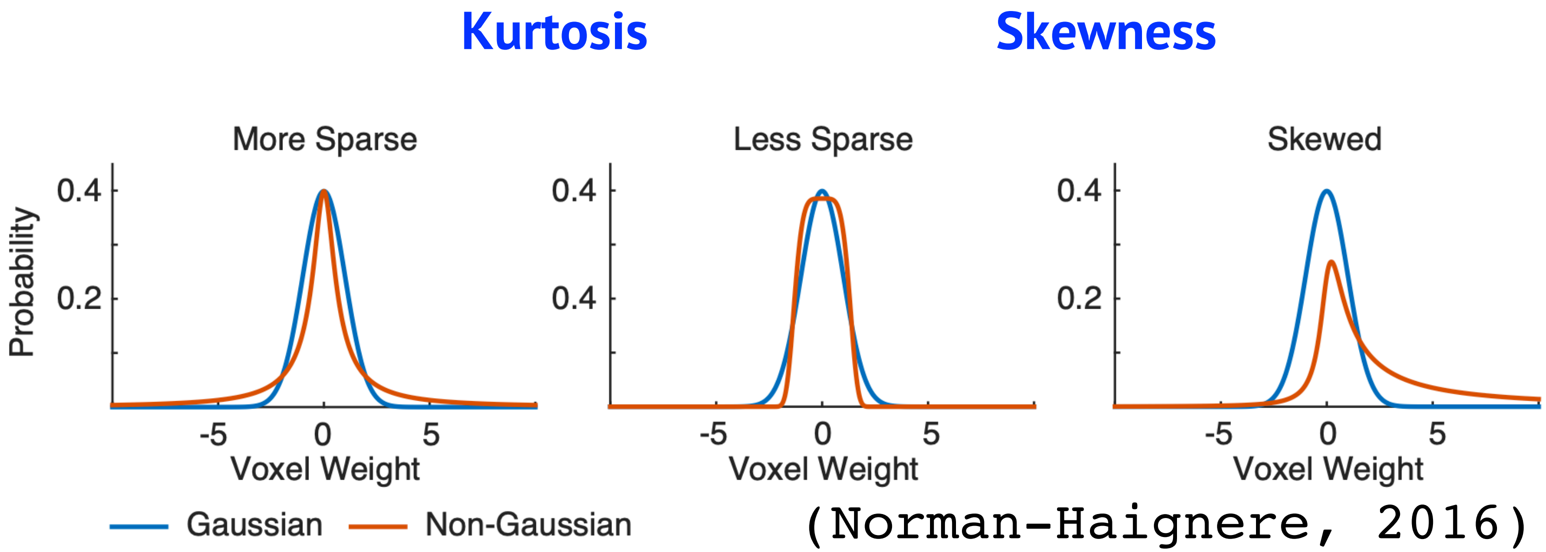
$$x_{ij} = a_{i1}s_{1j} + \cdots + a_{in}\, s_{nj} = \sum_{k=1}^{n} a_{ik}s_{kj}$$

- **Assumption #3**: Non-Gaussian distribution of true sources

ICA Assumption

- **Assumption #1**: (Statistical) Independent true sources
- **Assumption #2**: Truly linear mixing
- **Assumption #3**: Non-Gaussian distribution of true sources

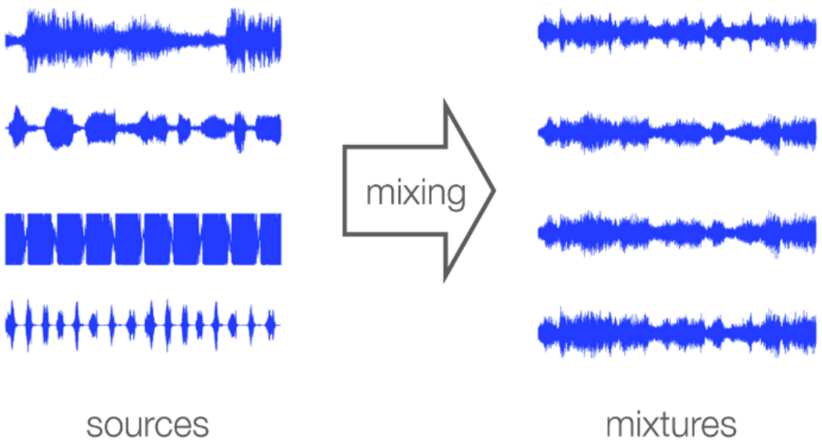




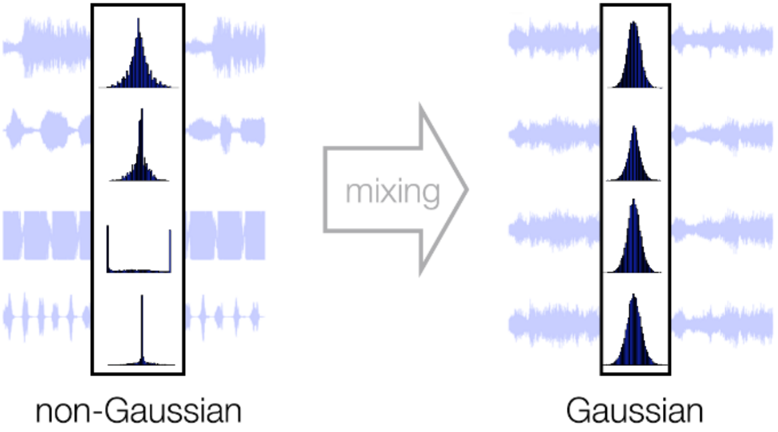
(Norman-Haignere, 2016)

ICA Assumption

- **Assumption #1**: (Statistical) Independent true sources
- **Assumption #2**: Truly linear mixing
- **Assumption #3**: Non-Gaussian distribution of true sources
  - Mixtures of sources have Gaussian distributions (**Central Limit Theorem; CLT**)



S $\qquad\qquad$ $AS = X$



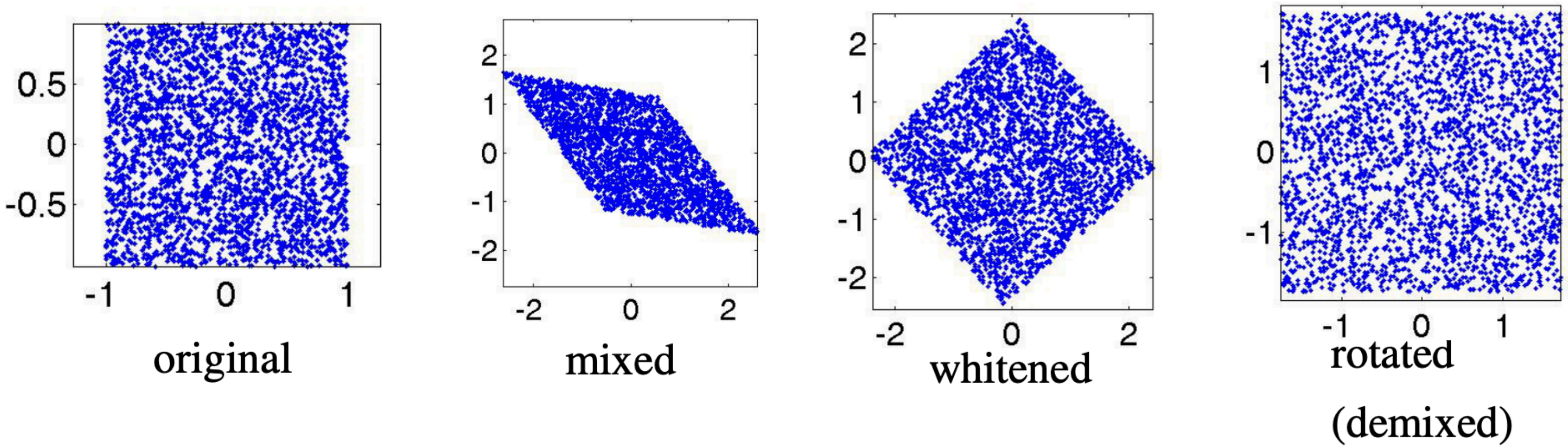
(FSL Course)

The Idea of ICA

- **Two Steps**

1. "Whiten" data to **remove co/variance.**
2. Rotate whitened PCA components to **maximize non-Gaussianity.**

We can use PCA for whitening!



(Póczos, 2009)

The Idea of ICA

- **Two Steps**

1. "Whiten" data to **remove co/variance.**

We can use PCA for whitening!

2. Rotate whitened PCA components to **maximize non-Gaussianity.**

   = to minimize Mutual information(KL Divergence)

   = to maximize Likelihood

Minimization of Mutual Information

Maximization of Non-Gaussianity

Maximization of Likelihood

$$I(Y_1, Y_2) = \sum_{i=1}^{2} H(Y_i) - H(Y_1, Y_2)$$

Mutual Information

Marginal Entropy

Joint Entropy

The Idea of ICA

Goal: Find $\boldsymbol{W}$ such that $(\boldsymbol{W}^{\boldsymbol{T}})^{-\mathbf{1}} = \boldsymbol{A}$

- **Measurement of Non-Gaussianity**
  - Negentropy (Gold standard)
  - Kurtosis (approximation)
  - Skew (approximation)

(Norman-Haignere, 2016)

The Idea of ICA

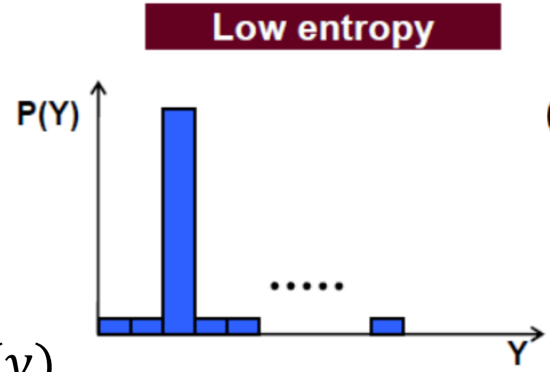
Goal: Find $\boldsymbol{W}$ such that $(\boldsymbol{W}^{\boldsymbol{T}})^{-1} = \boldsymbol{A}$

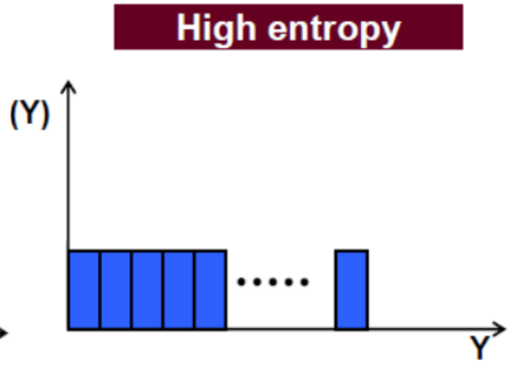
■ Measurement of Non-Gaussianity

- Negentropy (Gold standard)
  - Definition: difference in entropy from a Gaussian

$$J(y) = H(y_{gauss}) - H(y)$$

  - Gaussian distribution is maximally entropic (for fixed variance)
  - Maximizing negentropy closely related to minimizing mutual information
  - Cons: in practice can be hard to measure and optimize





- Kurtosis (approximation)
- Skew (approximation)

(Norman-Haignere, 2016)

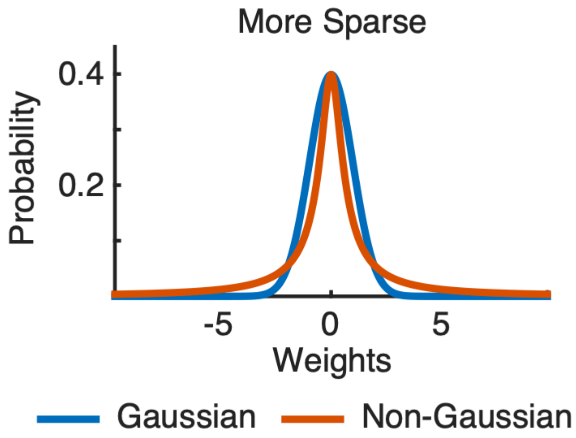
The Idea of ICA

Goal: Find $\boldsymbol{W}$ such that $(\boldsymbol{W}^{\boldsymbol{T}})^{-\mathbf{1}} = \boldsymbol{A}$

- **Measurement of Non-Gaussianity**
  - Negentropy (Gold standard)
  - Kurtosis (approximation)
    - Definition: 4th moment of the distribution

$$\mathbb{E}[y^4]$$

    - Useful for sparse, 'heavy tailed' distributions (which are common)
    - Very easy to measure and optimize
    - Cons: only useful if the source distributions are sparse, sensitive to outliers
  - Skew (approximation)



(Norman-Haignere, 2016)

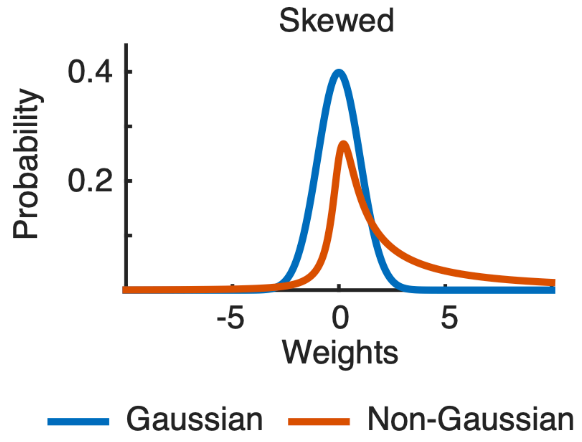
The Idea of ICA

Goal: Find $\boldsymbol{W}$ such that $(\boldsymbol{W}^{\boldsymbol{T}})^{-1} = \boldsymbol{A}$

- **Measurement of Non-Gaussianity**
  - Negentropy (Gold standard)
  - Kurtosis (approximation)
  - Skewness (approximation)
    - Definition: 3th moment of the distribution

$$\mathbb{E}[y^3]$$

    - Useful for distributions with a single heavy tail
    - Again easy to measure and optimize
    - Cons: only useful if the source distributions are skewed



(Norman-Haignere, 2016)

The Idea of ICA

Goal: Find $\boldsymbol{W}$ such that $(\boldsymbol{W}^{\boldsymbol{T}})^{-\mathbf{1}} = \boldsymbol{A}$

- **Maximization of Non-Gaussianity measure**
  - Brute-force search
  - Gradient-based

(Norman-Haignere, 2016)

The Idea of ICA

Goal: Find $\boldsymbol{W}$ such that $(\boldsymbol{W}^{\boldsymbol{T}})^{-\mathbf{1}} = \boldsymbol{A}$

- **Maximization of Non-Gaussianity measure**
  - Brute-force search
    - E.g., iteratively rotate pairs of components to maximize non-Gaussianity
    - Easy-to-implement, effective in low-dimensional spaces
  - Gradient-based

(Norman-Haignere, 2016)

The Idea of ICA

Goal: Find $\boldsymbol{W}$ such that $(\boldsymbol{W}^{\boldsymbol{T}})^{-\mathbf{1}} = \boldsymbol{A}$

- **Maximization of Non-Gaussianity measure**
  - Brute-force search
  - Gradient-based
    - More complicated to implement, effective in high dimensions

All optimization algorithms attempt to find local, not global, optima
- ➢ Useful to test stability of local optima
- ➢ E.g., run algorithm many times from random starting points

(Norman-Haignere, 2016)

The Idea of ICA

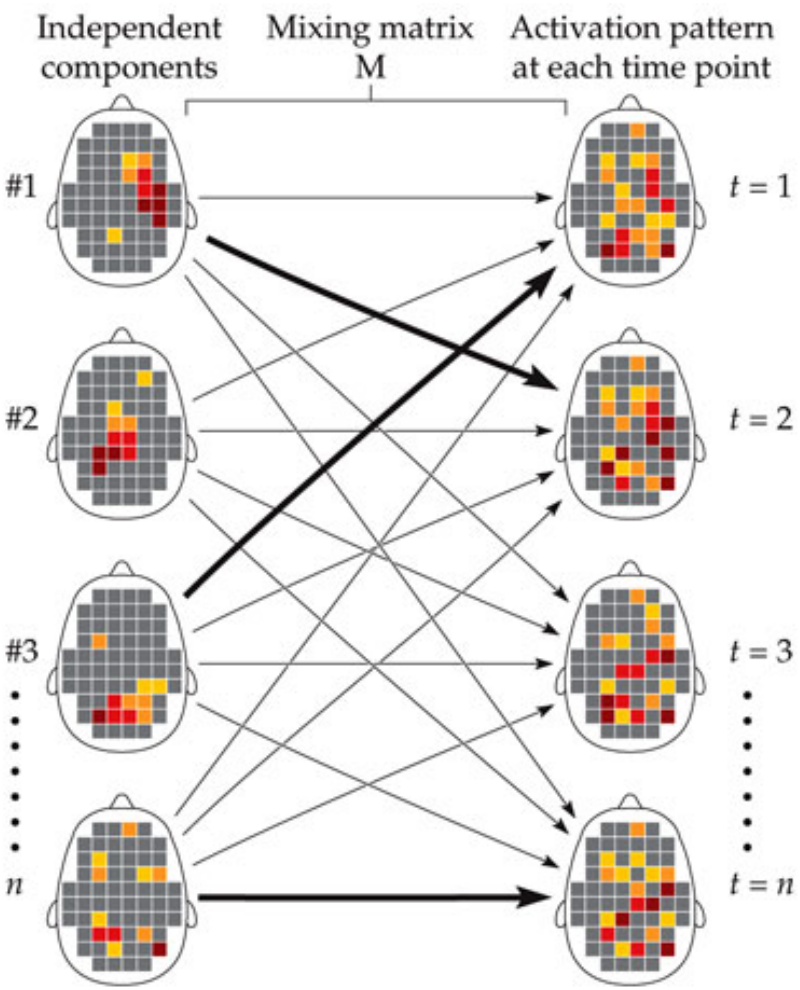
- **Algorithms**

There are more than 100 different ICA algorithms…

- Mutual information (MI) estimation
  - Kernel-ICA [Bach & Jordan, 2002]
- Entropy, negentropy estimation
  - Infomax ICA [Bell & Sejnowski 1995]
  - RADICAL [Learned-Miller & Fisher, 2003]
  - FastICA [Hyvarinen, 1999]
  - [Girolami & Fyfe 1997]
- ML estimation
  - KDICA [Chen, 2006]
  - EM-ICA [Welling]
  - [MacKay 1996; Pearlmutter & Parra 1996; Cardoso 1997]
- Higher order moments, cumulants based methods
  - JADE [Cardoso, 1993]
- Nonlinear correlation based methods
  - [Jutten and Herault, 1991]

(Póczos, 2009)

The Idea of ICA



Functional Magnetic Resonance Imaging 2e, Figure 11.2

(Huettel, Song & McCarthy, 2008)

The Idea of ICA

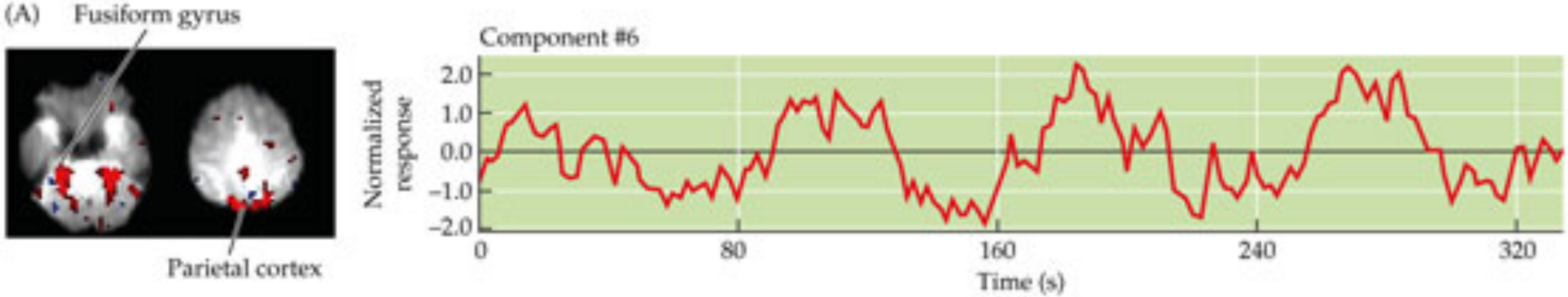
- **How many components?**
  - too many
    - splitting of components
    - hard to dig through
  - too few
    - clumping of components
  - some algorithms can estimate # components
- **How do you make sense of them?**
  - visual inspection
  - sorting

Sorting components

- variance accounted for by component
- spatial correlation with known areas
  - regions of interest (e.g., fusiform face area)
  - networks of interest (e.g., default mode network)
- temporal correlation with known events
  - task predictors

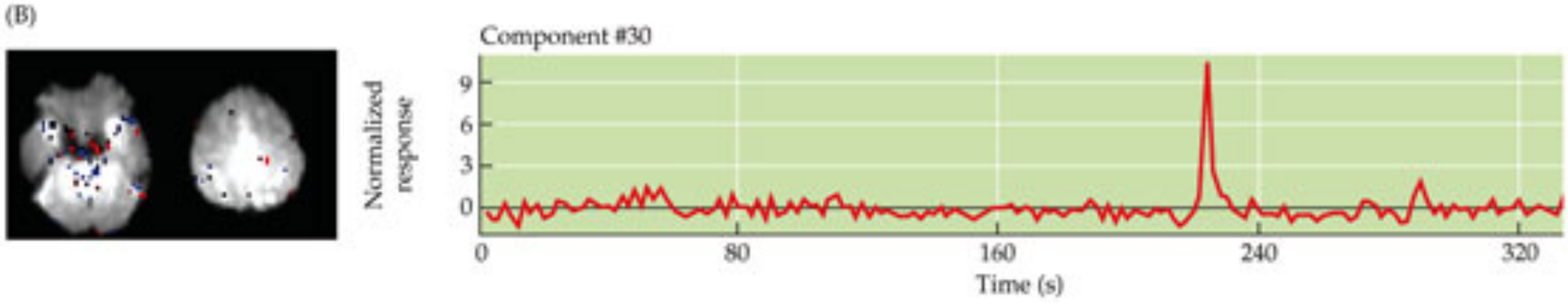
(NeWBI 4 fMRI)

The Idea of ICA



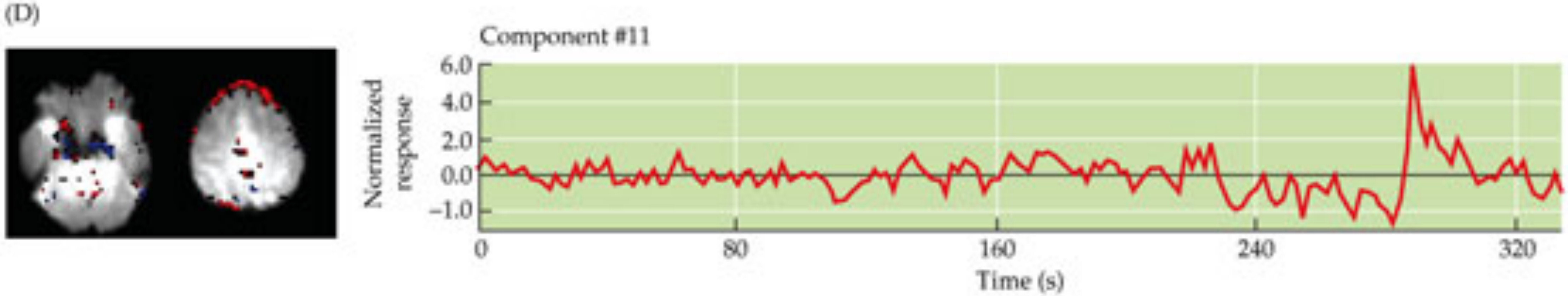
(Huettel, Song & McCarthy, 2008)

The Idea of ICA



(Huettel, Song & McCarthy, 2008)

The Idea of ICA



(Huettel, Song & McCarthy, 2008)

Interim Summary

■ **Assumptions**
- **Assumption #1**: (Statistical) Independent true sources
- **Assumption #2**: Truly linear mixing
- **Assumption #3**: Non-Gaussian distribution of true sources

■ **Two Steps**
1. "Whiten" data to **remove co/variance**.
2. Rotate whitened PCA components to **maximize non-Gaussianity.**

Goal: Find $\boldsymbol{W}$ such that $(\boldsymbol{W}^{\boldsymbol{T}})^{-1} = \boldsymbol{A}$

■ **Measurement of Non-Gaussianity**
- Negentropy (Gold standard)
- Kurtosis (approximation)
- Skew (approximation)

■ **Maximization of Non-Gaussianity measure**
- Brute-force search
- Gradient-based

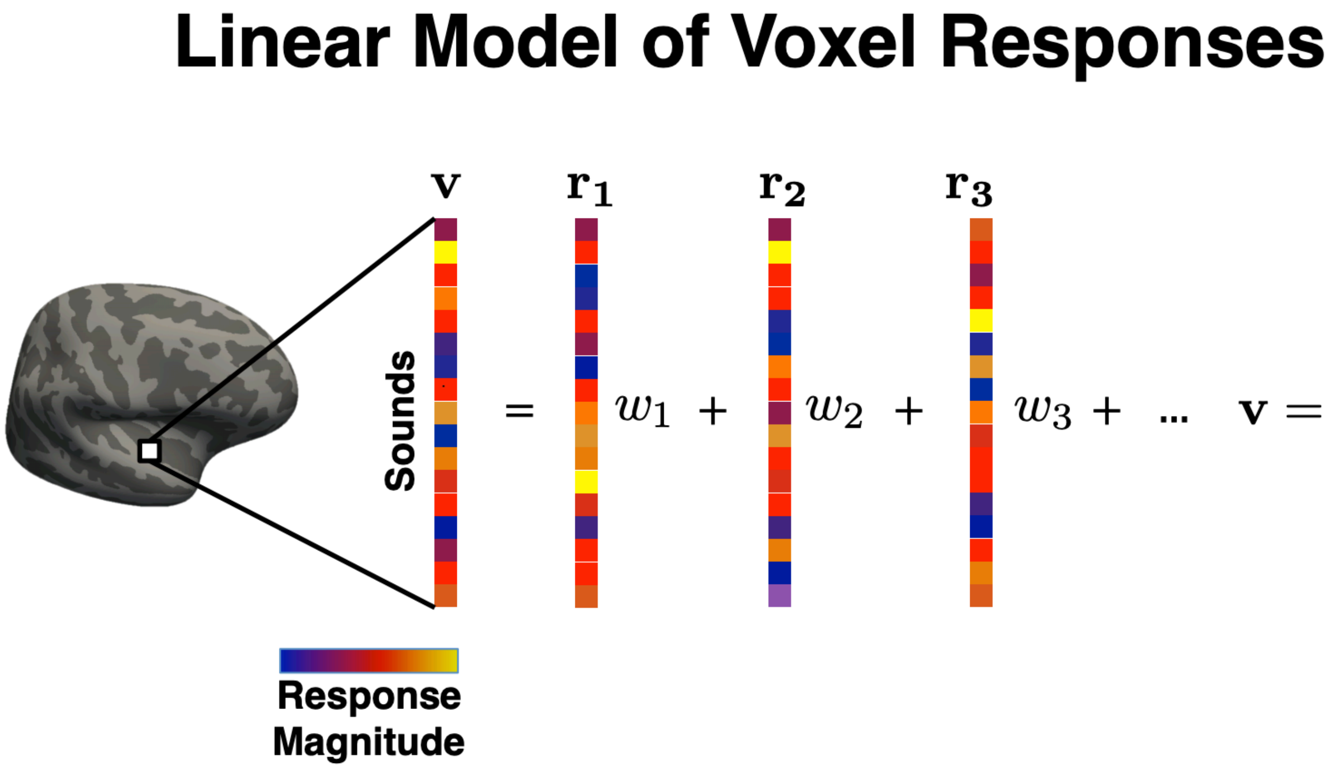
There are more than different 100 ICA algorithms…

■ **How many components?**
- too many
- too few
- some algorithms can estimate # components

■ **How do you make sense of them?**
- visual inspection
- sorting

ICA vs. PCA

- Similarities



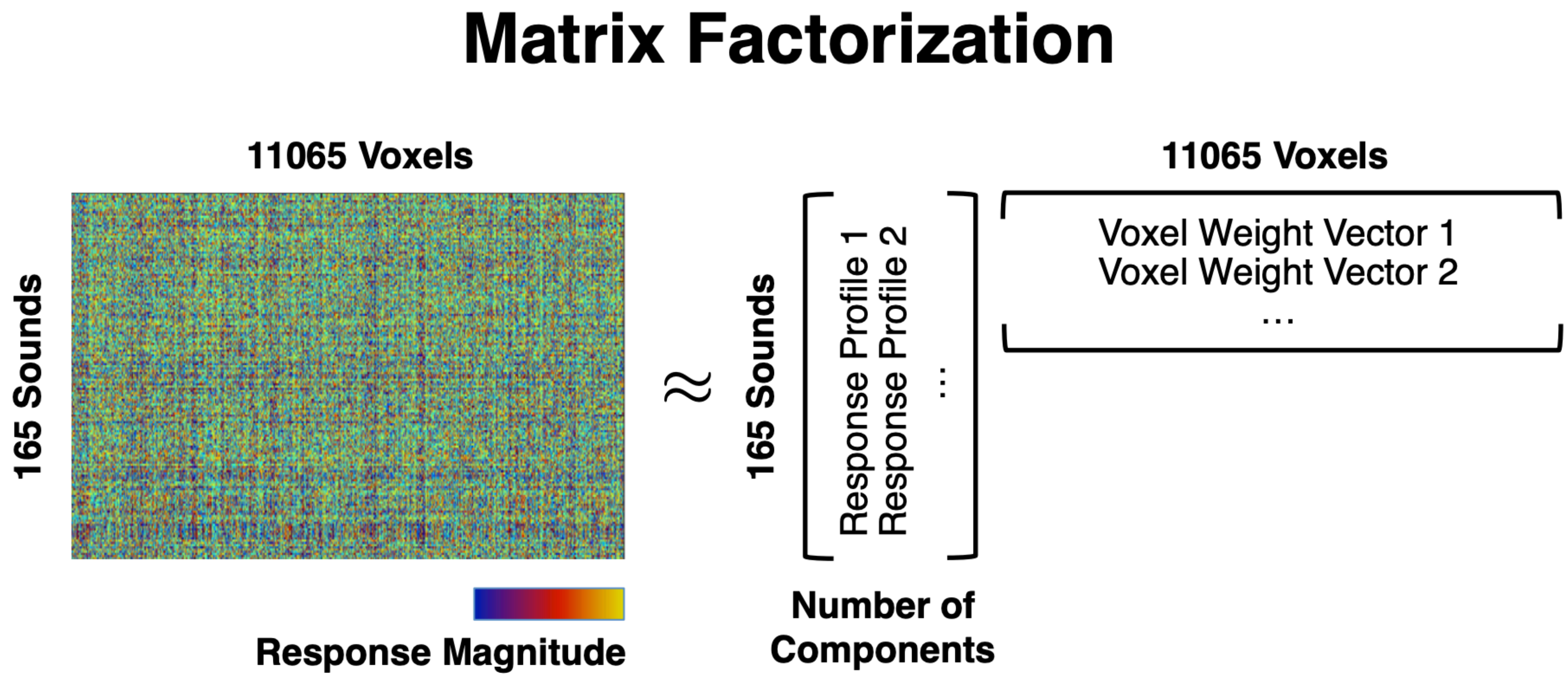
$$\mathbf{v} = \mathbf{r_1} w_1 + \mathbf{r_2} w_2 + \mathbf{r_3} w_3 + \ldots \quad \mathbf{v} = \sum_{i=1}^{N} \mathbf{r_i} w_i$$

Voxel responses modeled as weighted sum of response profiles

(Norman-Haignere, 2016)

ICA vs. PCA

- **Similarities**



(Norman-Haignere, 2016)

ICA vs. PCA

- **Differences**

ICA

- $X = AS$
- **Non-Gaussianity**
- ICA does **not** do compression
  - Same # of features (M=N)
- ICA just removes correlations, **and** higher order dependence (3rd and 4th orders)
- Components are **equally important**

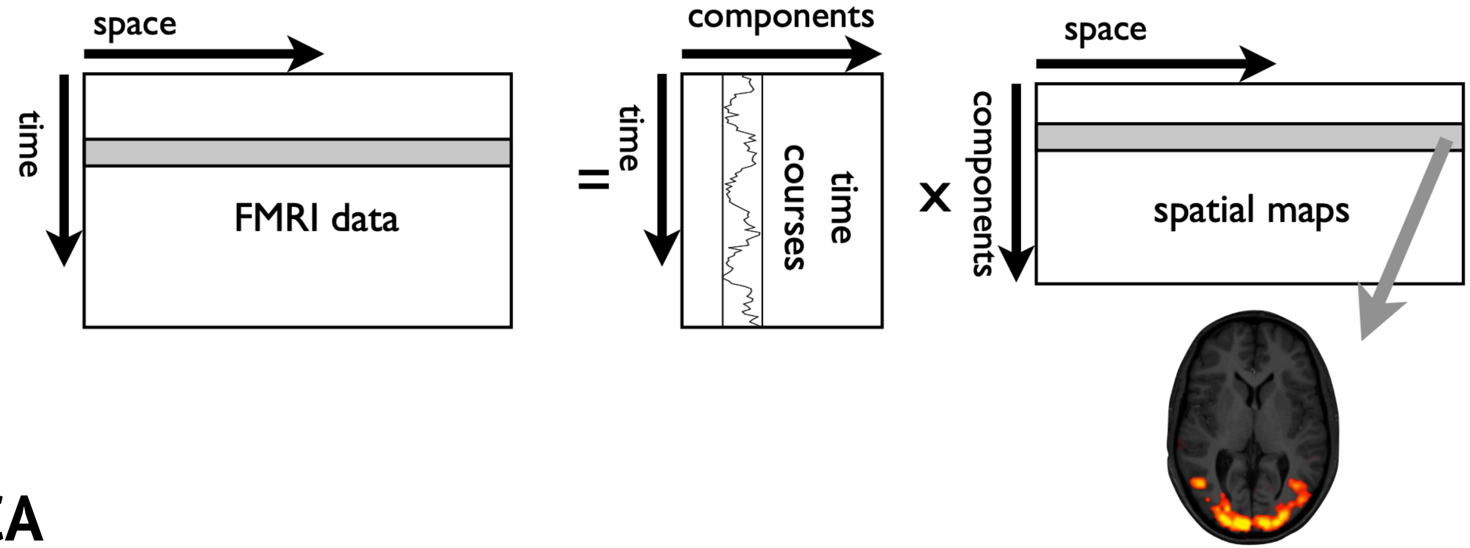
PCA

- $X = US \ (U^T U = I)$
- **Variance = Relevance**
- PCA does **compression**
  - M < N
- PCA just removes correlations, **not** higher order dependence (2nd order)
- Some components are **more important** than others (based on eigenvalues)

(Póczos, 2009)

ICA Types

- **Spatial ICA**
  - Independent components are estimated by maximizing independence in space (spatially independent) ⇨ common approach

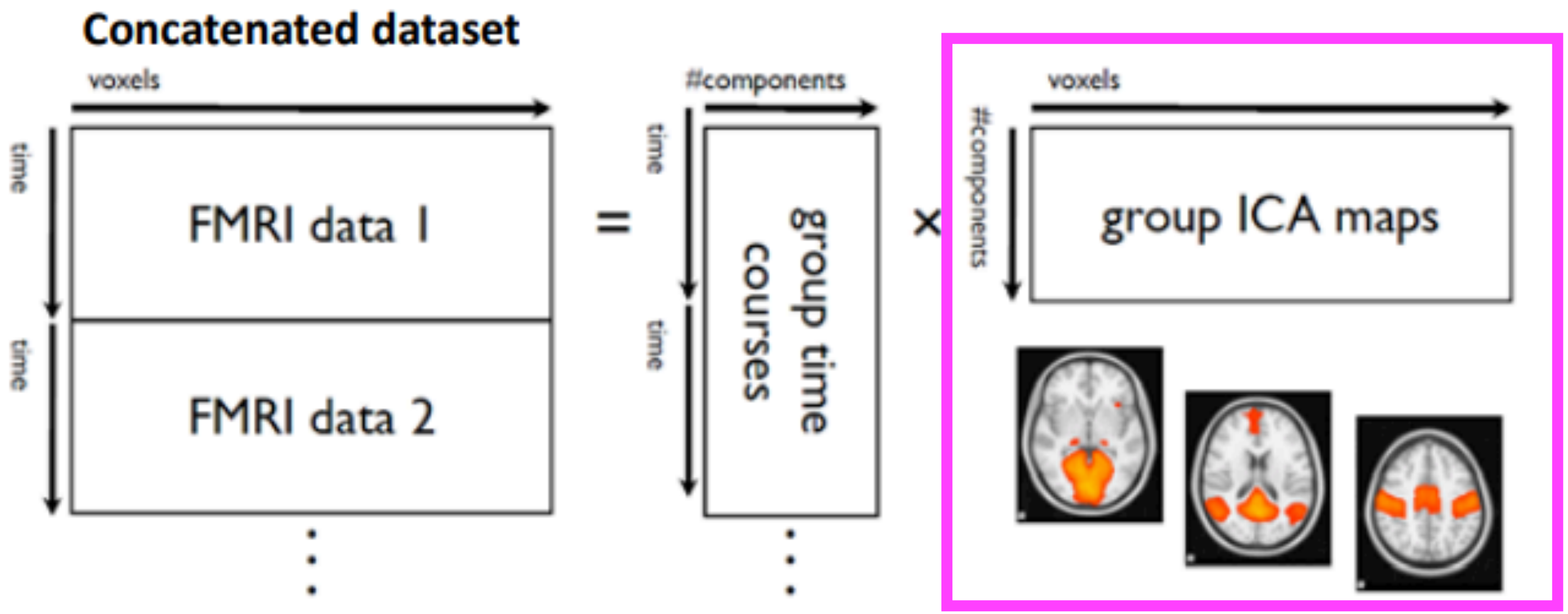


- **Temporal ICA**
  - Independent components are estimated by maximizing independence in time

(FSL Course)

Group ICA

- Concatenate all participants' data temporally/spatially, then run ICA.

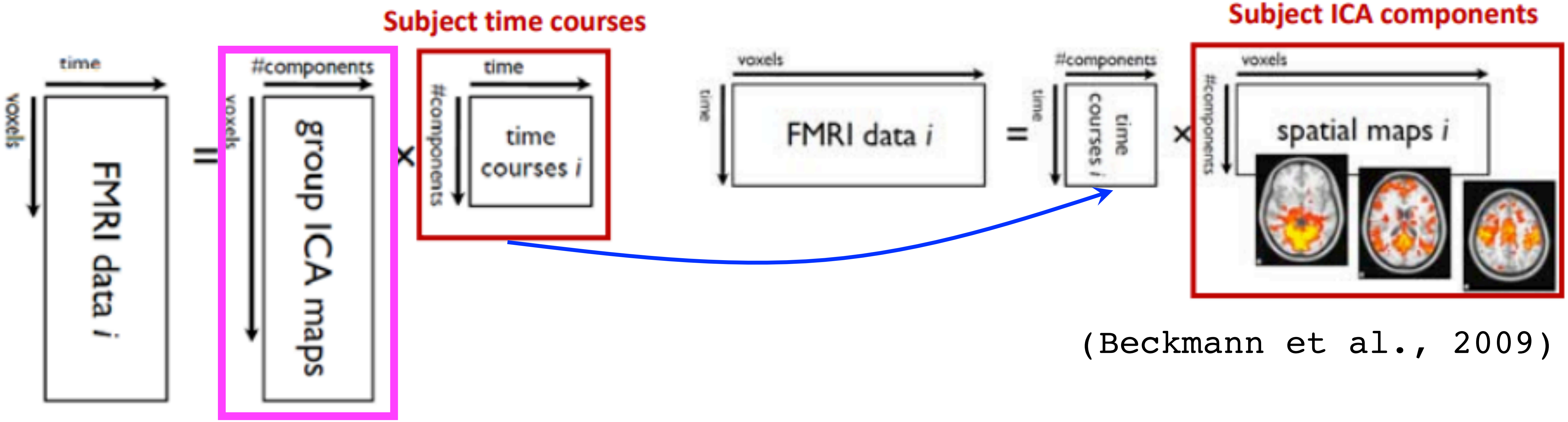


(Beckmann et al., 2009)

(Lu, 2019)

Reconstructed Subject ICA

- **Dual Regression**



(Beckmann et al., 2009)

1. Regress group maps into each subjects' 4D data to find subject- specific timecourses

2. Regress these timecourses back into the 4D data to find subject-specific spatial maps

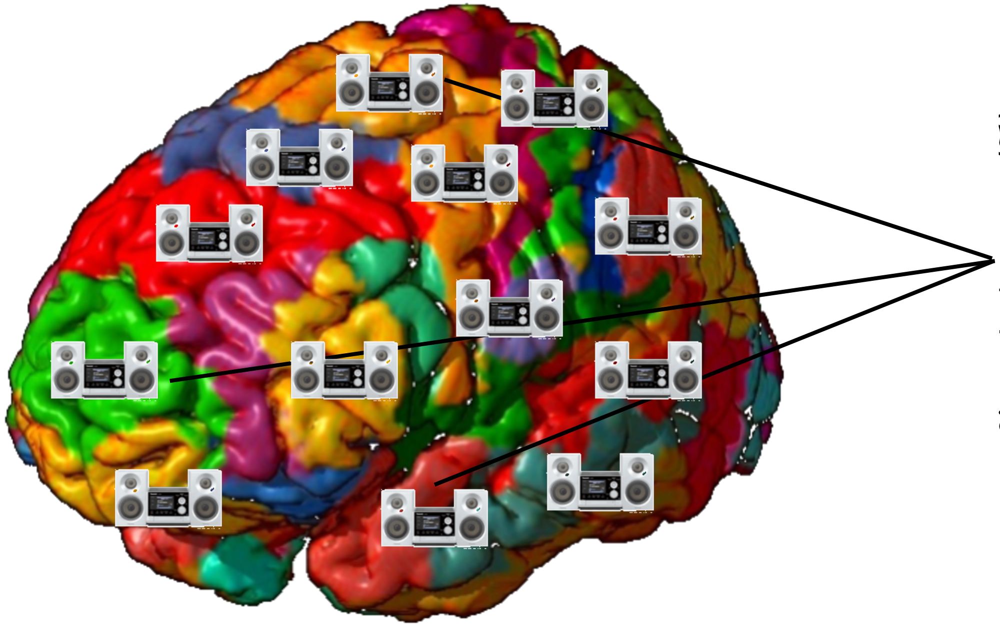
(Lu, 2019)

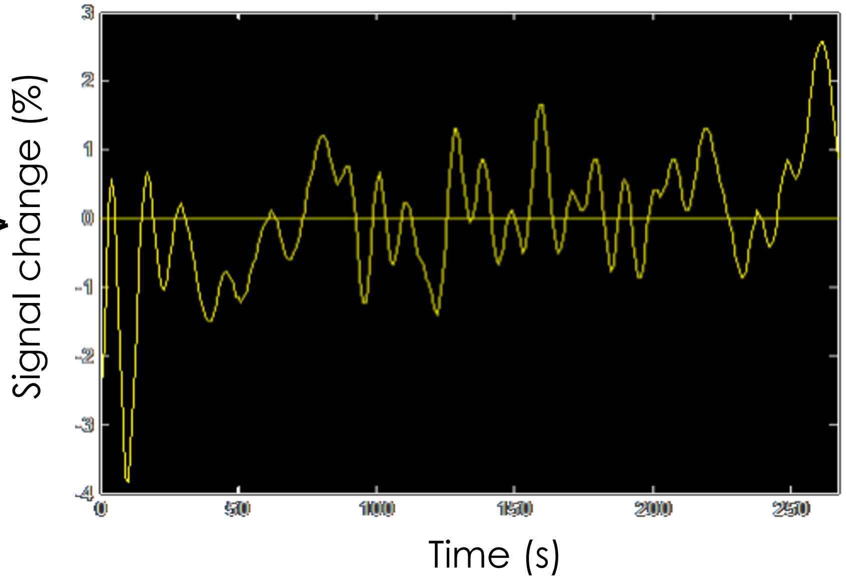
ICA Limitations

- **Assumption #1**: (Statistical) Independent true sources
- **Assumption #2**: Truly linear mixing
- **Assumption #3**: Non-Gaussian distribution of true sources

ICA Limitations

- **Assumption #1**: (Statistical) Independent true sources
- **Assumption #2**: Truly linear mixing





- **Assumption #3**: Non-Gaussian distribution of true sources

(NeWBI 4 fMRI)

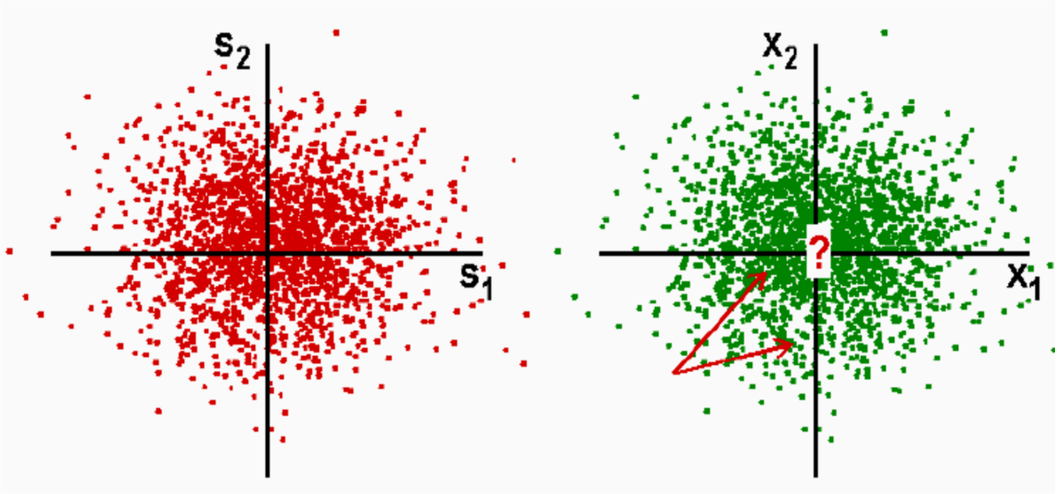
ICA Limitations

- **Assumption #1**: (Statistical) Independent true sources
- **Assumption #2**: Truly linear mixing
- **Assumption #3**: Non-Gaussian distribution of true sources

Negentropy (Gold standard)

- Mutual information (MI) estimation
  - Kernel-ICA [Bach & Jordan, 2002]
- Entropy, negentropy estimation
  - Infomax ICA [Bell & Sejnowski 1995]
  - RADICAL [Learned-Miller & Fisher, 2003]
  - FastICA [Hyvarinen, 1999]
  - [Girolami & Fyfe 1997]
- ML estimation
  - KDICA [Chen, 2006]
  - EM-ICA [Welling]
  - [MacKay 1996; Pearlmutter & Parra 1996; Cardoso 1997]
- Higher order moments, cumulants based methods
  - JADE [Cardoso, 1993]
- Nonlinear correlation based methods
  - [Jutten and Herault, 1991]

(Póczos, 2009)

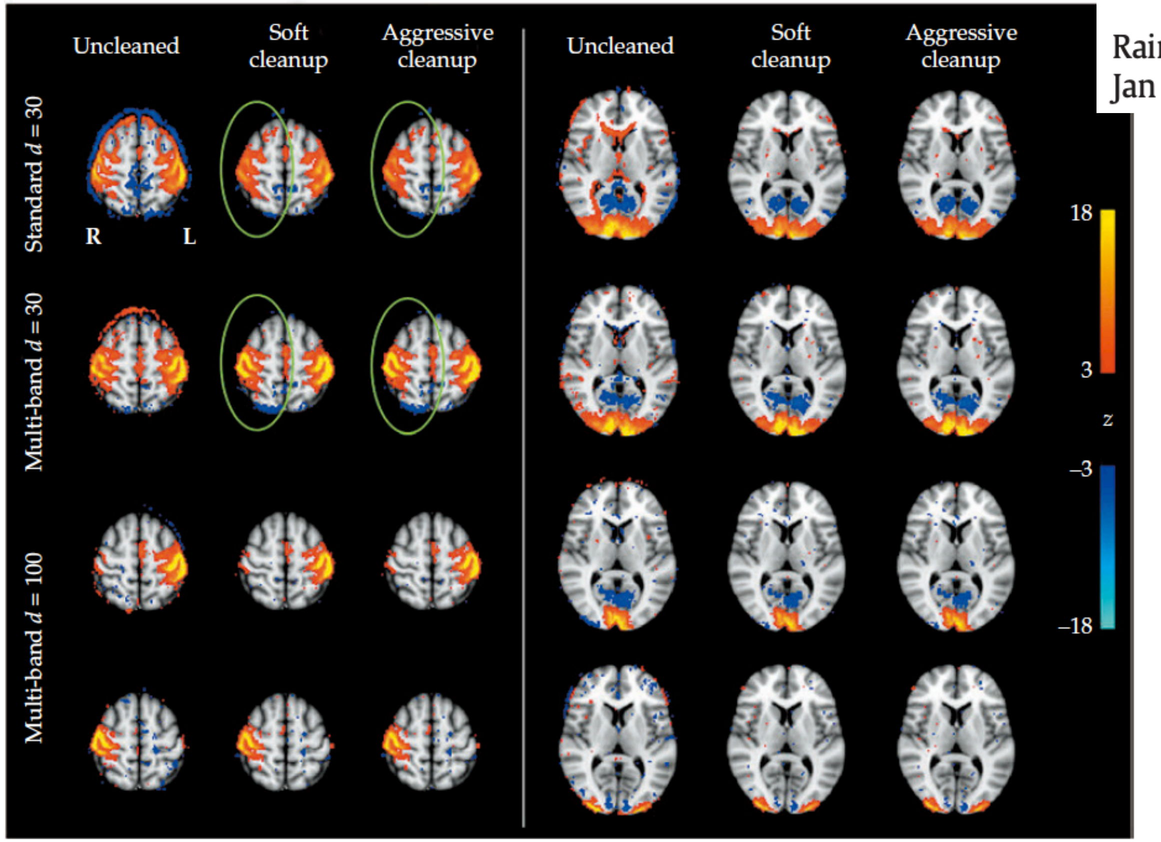


Application

- **Data denoising**
- **Network Identification (functional connectivity)**

Application

- **Data denoising**



ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data

Raimon H.R. Pruim [a,b,*], Maarten Mennes [a,b], Daan van Rooij [b,c], Alberto Llera [b], Jan K. Buitelaar [a,b,d], Christian F. Beckmann [a,b,e]

(Pruim et al., 2015, Neuroimage)

PART 5 (b9-b10)

Functional images
Smoothing and ICA-AROMA

b9: smoothing
- smoothing functional images with the FWHM 5 mm smoothing kernel.

b10: ICA-AROMA
- A data-driven method to identify and remove motion-related independent components from functional MRI data.
- https://github.com/rhr-pruim/ICA-AROMA

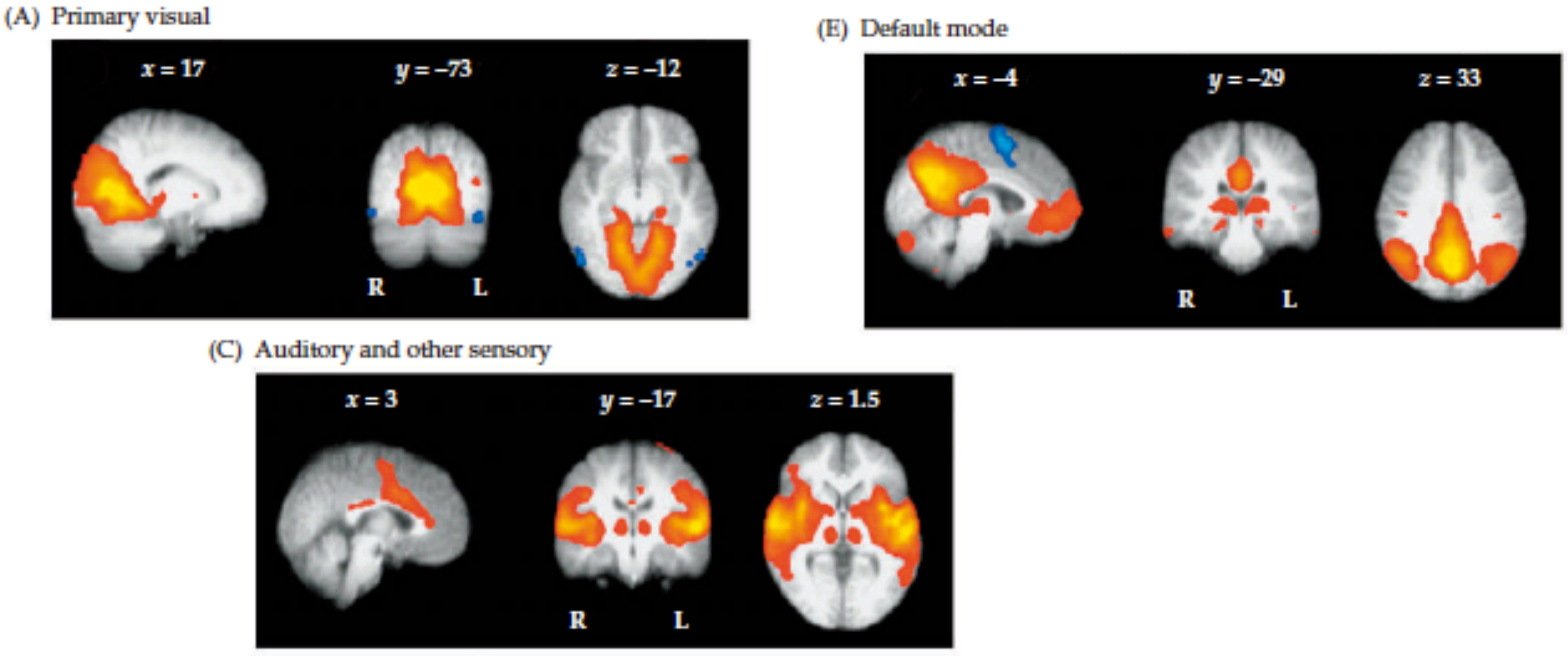
Cocoanlab's fMRI data preprocessing pipeline (Link)

- Network Identification (functional connectivity)

(Huettel, Song & McCarthy, 2008)

Application

- ○ Data denoising
- ○ **Network Identification (functional connectivity)**



(Huettel, Song & McCarthy, 2008)

Conclusion

- For ICA to infer underlying components, they must have non-Gaussian and statistically independent sources (voxel weights for fMRI).

- When the assumptions hold, the method is very effective, and requires few additional assumptions about the nature of the underlying sources.

- While ICA is a powerful technique with few assumptions on the nature of the observations and the underlying sources, it must be remembered that ICA does have some intrinsic limitations.

- In fMRI data, ICA has showed quite good performance in both data denoising and network identification.

Reference #1

- NeWBI 4 fMRI. Independent Component Analysis (ICA) [PowerPoint slides]. http://www.newbi4fmri.com/
- Mike X Cohen. PCA & multivariate signal processing, applied to neural data. *Udemy*. https://www.udemy.com/course/dimension-reduction-and-source-separation-in-neuroscience/
- Norman-Haignere, S. (2016). Dimensionality Reduction II: ICA. Center For Brains, Minds, and Machines. https://cbmm.mit.edu/learning-hub/tutorials/computational-tutorial/dimensionality-reduction-ii
- Huettel, S. A., Song, A. W., & McCarthy, G. (2009). Functional magnetic resonance imaging. *Sunderland, Mass: Sinauer Associates*.
- FSL Course. Resting State: Independent Component Analysis [PowerPoint slides]. Section 6: Resting-State fMRI: ICA and Dual Regression. http://fsl.fmrib.ox.ac.uk/fslcourse/lectures/Rest_E2.pdf

Reference #2

- Póczos, B. (2009). Introduction to Independent Component Analysis [PowerPoint slides]. *Machine Learning Department, Carnegie Mellon University*. http://www.cs.cmu.edu/~./bapoczos/other_presentations/ICA_26_10_2009.pdf
- Lu, CF. (2019). Brain Network– Independent component analysis [PowerPoint slides]. *Department of Biomedical Imaging and Radiological Sciences, National Yang-Ming University*. https://www.ym.edu.tw/~cflu/fMRIanalysis_Class10_CFLu.pdf