University of Southampton

isvr Institute of Sound and Vibration Research

Independent Component Analysis:
Fundamentals and applications to biomedicine

Christopher J. James

Outline

- Introduction
- When can ICA be used?
- Algorithms
- Performance
- Applications to Biomedicine
- Practical Difficulties
- Summary

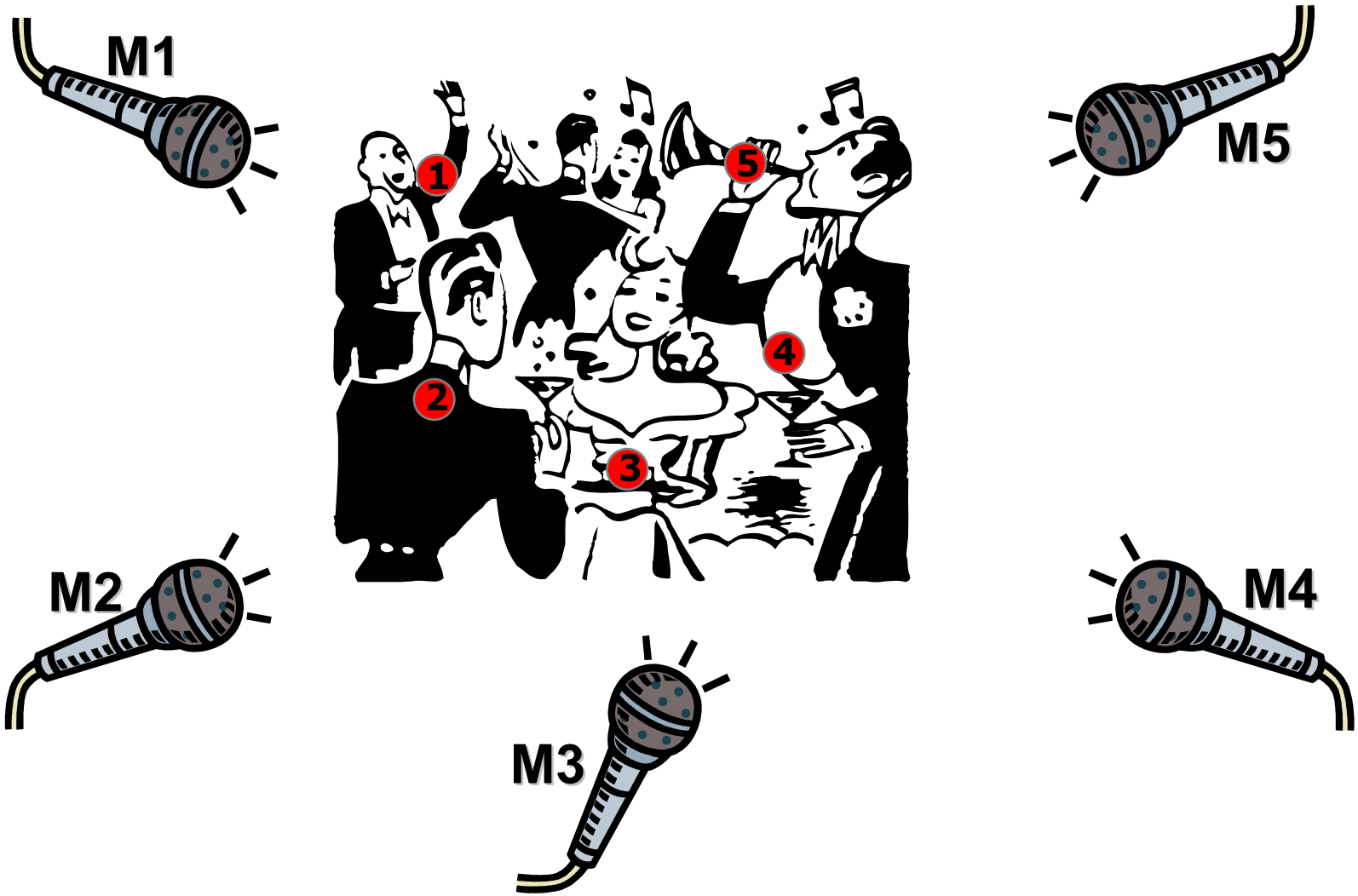
Introduction

- Independent Component Analysis (ICA) is a form of Blind Source Separation (BSS) and is used in the identification and separation of mixtures of sources with little prior information.

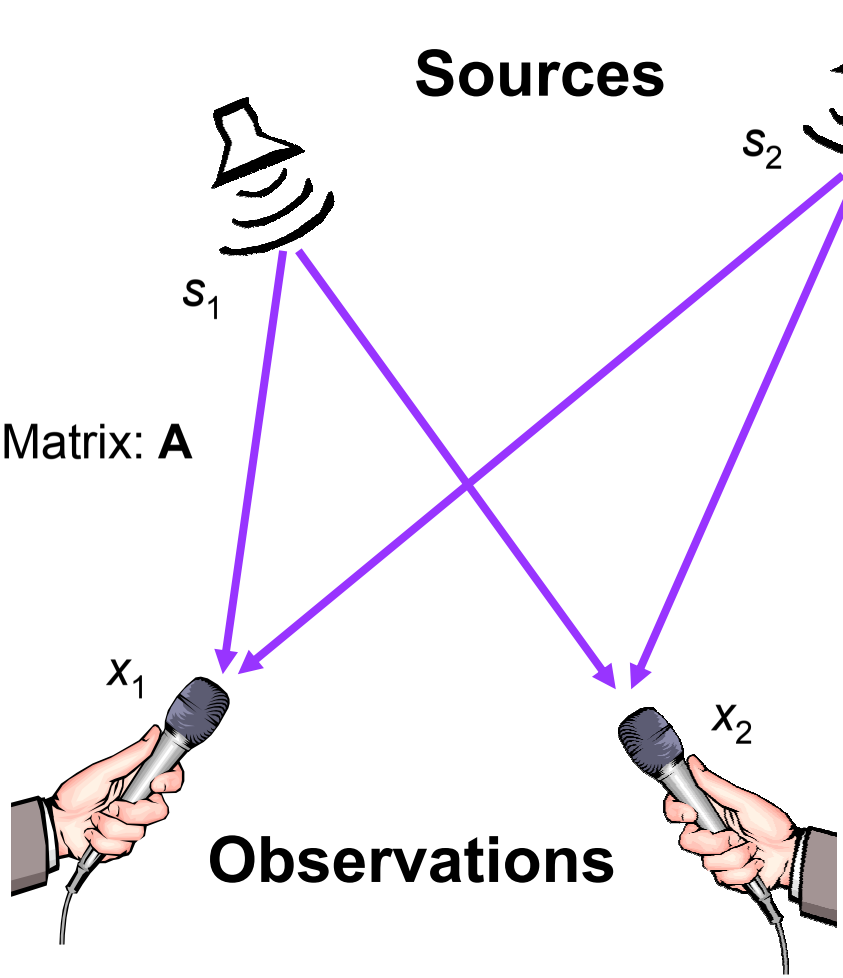
- Applications include:
  - Biomedical data
  - Audio signals
  - Financial data
  - Coding
  - Array processing

  … and most applications where traditionally Factor Analysis and PCA is currently used.

The ‘Cocktail Party’ problem



The ‘Cocktail Party’ problem



Mixing Matrix: **A**

x=As + e

(where **e** represents noise)

n sources

m observations

Assume *n*=*m*

The ICA model

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \forall \mathbf{x}, \mathbf{s} \in \Re^n$$

- Assuming **no-noise** and number of sources (*n*)=number of observations (*m*) [**square-mixing**], we can aim to find the inverse mixing matrix $\mathbf{W}=\mathbf{A}^{-1}$ to give estimates of the sources:

$$\hat{\mathbf{s}}(t) = \mathbf{W}\mathbf{x}(t)$$

- ICA achieves this by assuming the sources to be *statistically independent*, i.e.

$$p(\mathbf{s}) = \prod_i p(s_i)$$

Ambiguities with the ICA solution

- Permutation:
  - No ordering of the sources is specified such that the following is also a solution:

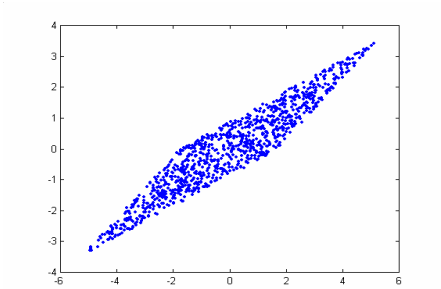
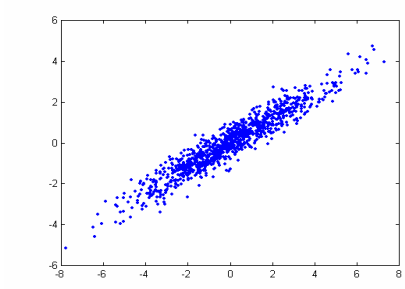
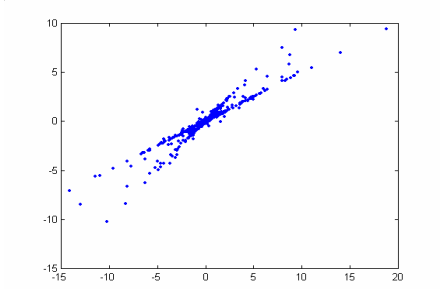
$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\mathbf{W}$$

- Scale:
  - No prior knowledge of scale is imposed so the following is also a solution:

$$\begin{bmatrix} \alpha & 1 \\ 1 & \beta \end{bmatrix}\mathbf{W}$$

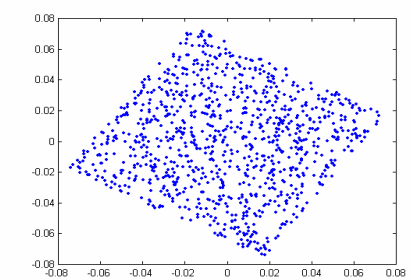
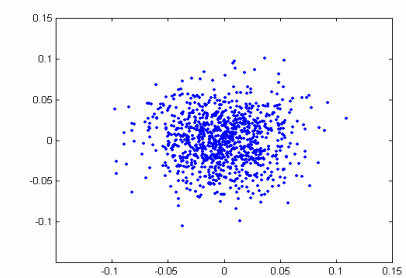
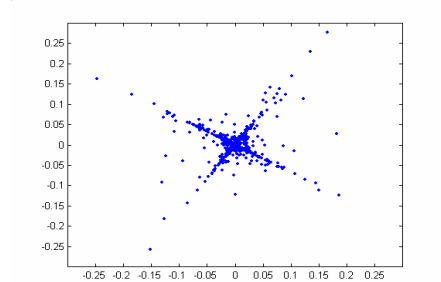
- These are, however, rather trivial and easily overcome

ICA: More than just Decorrelation

Kurtosis

-ve ← 0 → +ve

All decorrelated (PCA), not necessarily independent

HOS - nonGaussianity

- The previous example implies that as long as the sources are not Gaussian it should be possible to perform BSS. Comon (1994) states:

  Theorem: *For a vector **s** of independent entries with at most one Gaussian entry and for any invertible matrix **WA**, if the entries of **y=WAs** are independent then **y** is a copy of **s** (up to permutation and rescaling).*

- This implies that ICA is possible (for non-Gaussian sources) and that maximising independence provides a means of performing BSS.

Methods of Source Separation

- There are various ways of practically measuring independence:
  - **Higher Order Statistics:** After pre-whitening, maximising the absolute of the kurtosis finds independent directions.
  - **Mutual Information & Maximum Likelihood:** If $p_i(s_i)$ is assumed to be known *a priori*, ML is equivalent to nonlinear decorrelation.
  - **Time or Time-frequency Decorrelation:** Decorrelating the measured signals over different time-lags or different time-frequency decompositions.

Fast ICA

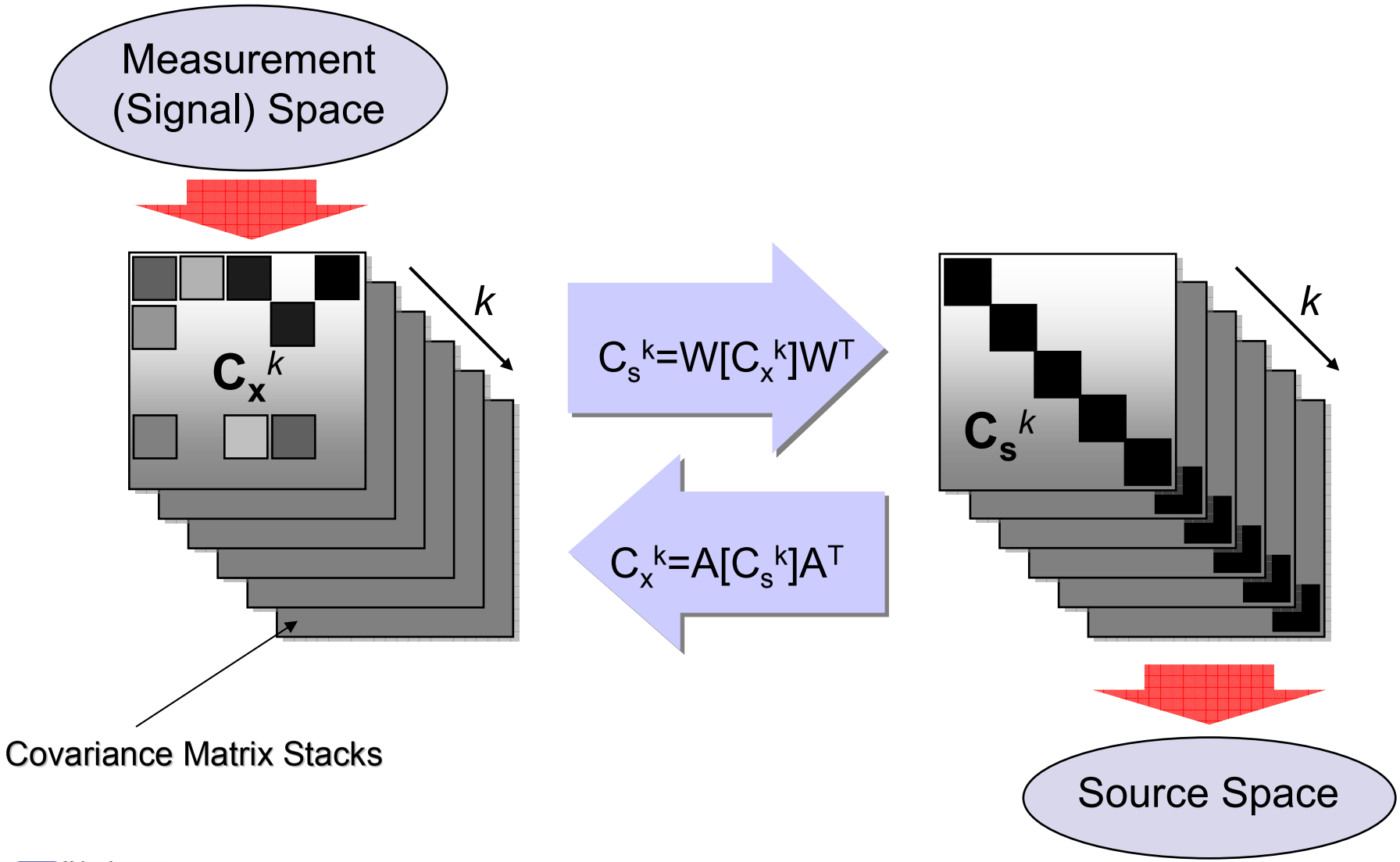
- FastICA attempts to separate underlying sources based on their 'non-Gaussianity'. The simple premise behind FastICA is that the fast fixed-point iterative algorithm undertakes to find projections that maximize the non-Gaussianity of components by their Kurtosis (the 4th order cumulant given to a random variable).
- As Kurtosis is identically zero for Gaussian distributed signals, the aim is to maximise the magnitude of the Kurtosis to make the estimated sources as *non-Gaussian* (and hence as *independent*) as possible.
- Here, the ICA problem is posed as an optimisation problem with the sources as its solution. The Kurtosis that is used to describe the peakedness of a distribution is defined as

$$kurt(\mathbf{x}) = E\{\mathbf{x}^4\} - 3\left(E\{\mathbf{x}^2\}\right)^2$$

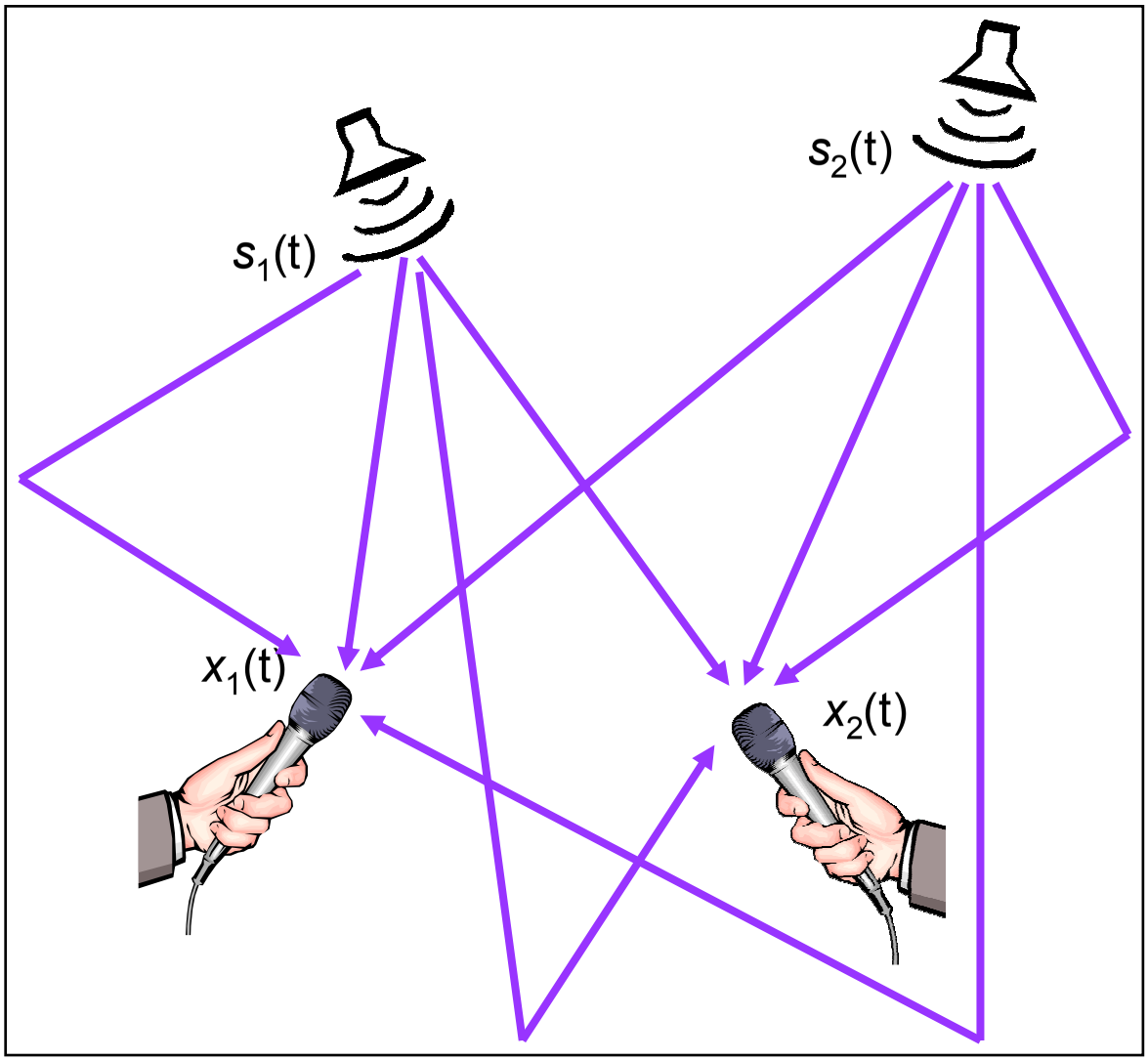
The Matrix Stack Diagonalisation Approach

- Performs joint diagonalisation of temporal cross-covariances matrices @ different lags (or delays)
- Cross-covariance data matrix (lag *k*):
  - $\mathbf{C}_x^k = \mathbf{A}[\mathbf{C}_s^k]\mathbf{A}^T$
- Cross-covariance source matrix (lag *k*):
  - $\mathbf{C}_s^k = \mathbf{W}[\mathbf{C}_x^k]\mathbf{W}^T$
- Many algorithms exist – such as LSDIAG_{TD}

The Matrix Stack Diagonalisation



Convolutive Mixtures



Convolutive mixing: **A(t)**

$$\mathbf{x}(t) = \sum_{\tau} \mathbf{A}(\tau)\mathbf{s}(t-\tau) + \mathbf{e}(t)$$

or

$$\mathbf{X}(\omega) = \mathbf{A}(\omega)\mathbf{S}(\omega) + \mathbf{E}(\omega)$$

In such audio applications it becomes beneficial to perform ICA in the frequency domain.

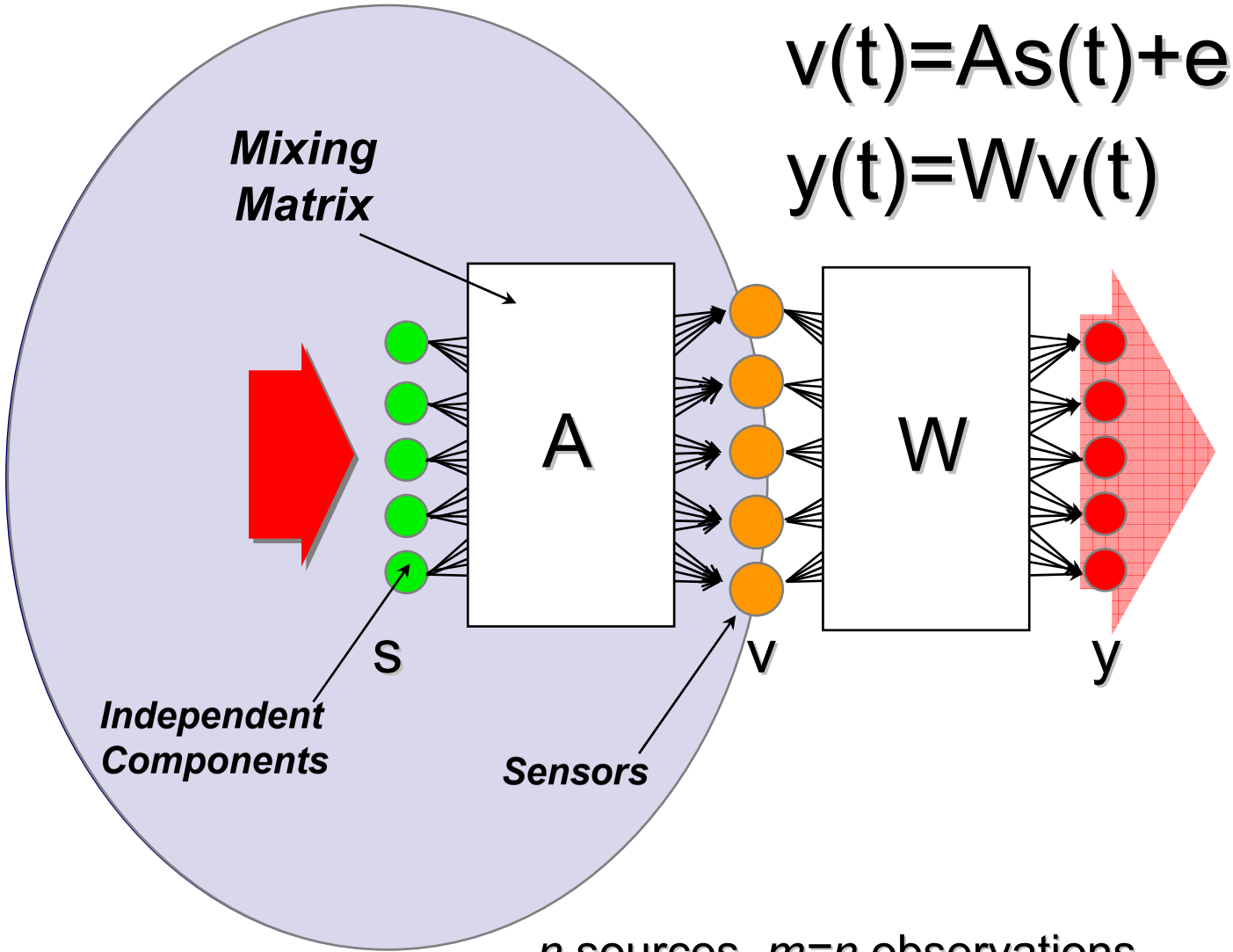
Applications to Biomedical Signals

- To separate physiological from non-physiological signals (artifact rejection)
- To separate physiological signals of interest
- EEG/MEG: normal vs abnormal patterns
- TaFECG: fetal ECG + Maternal ECG
- Tracking changing physiological sources over time

ICA in Biomedicine

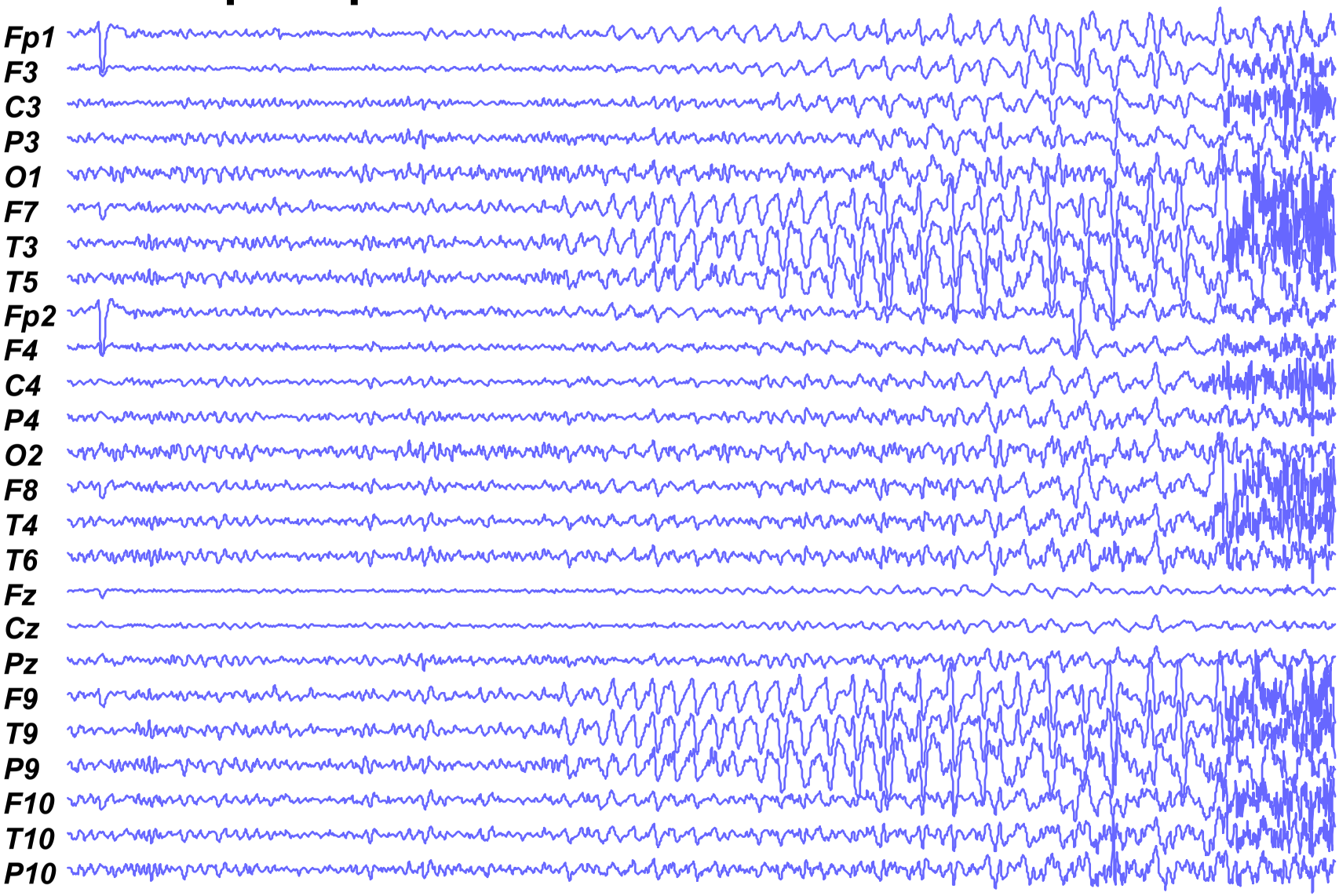
$$v(t)=As(t)+e(t)$$

$$y(t)=Wv(t)$$

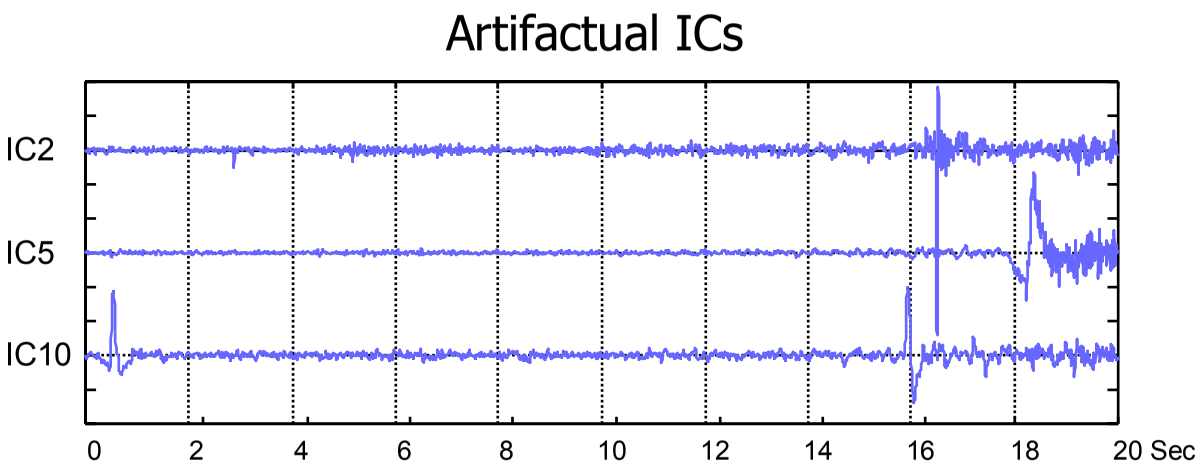


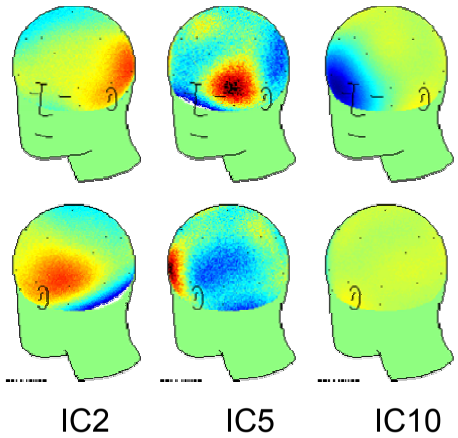
n sources, *m*=*n* observations

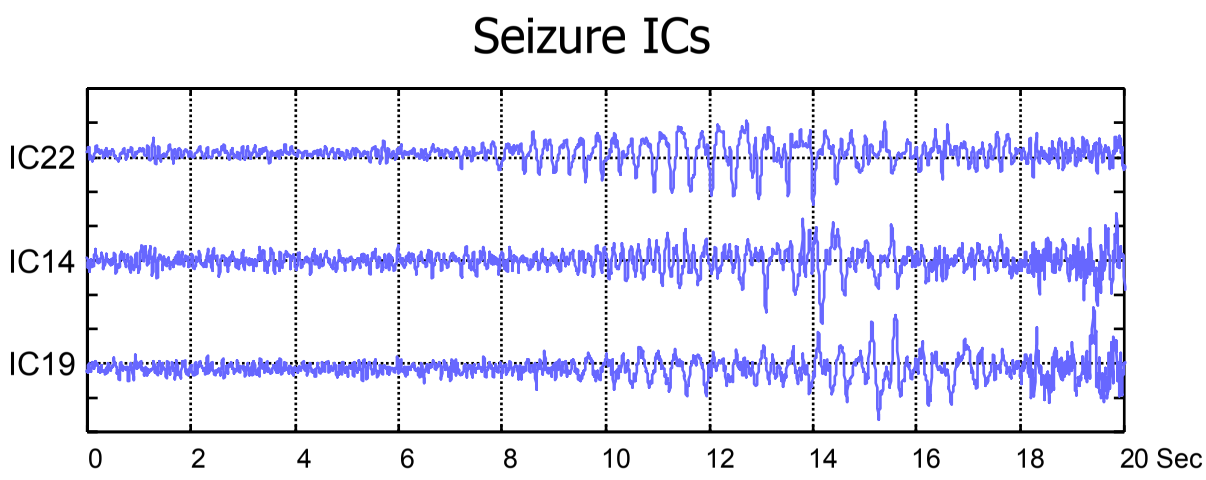
The Epileptic EEG

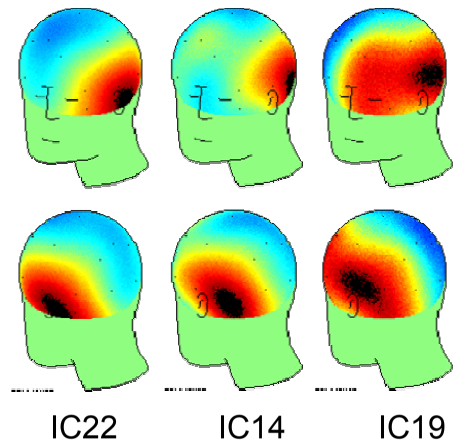


ICA and Seizure EEG





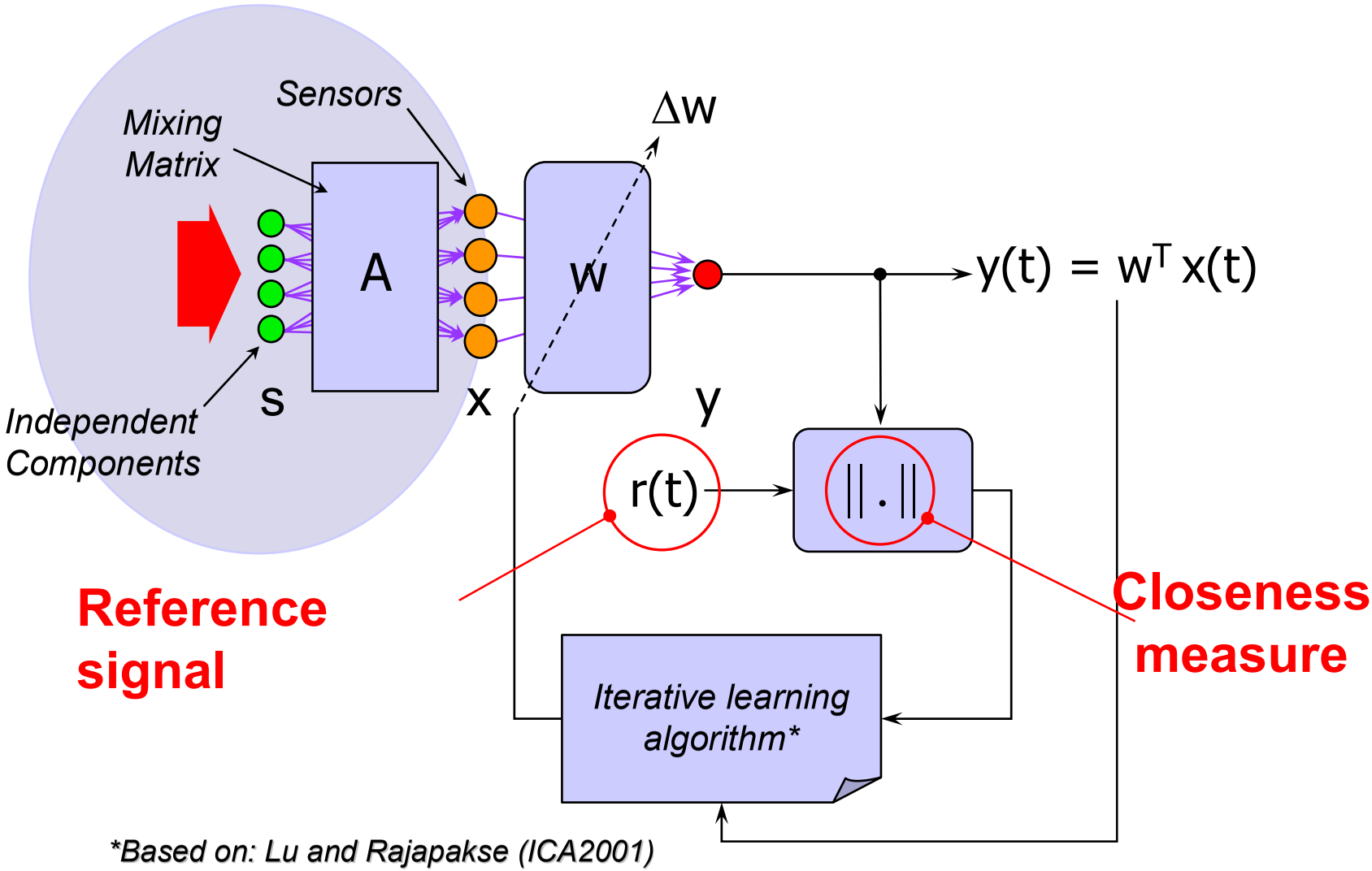




Extensions to ICA

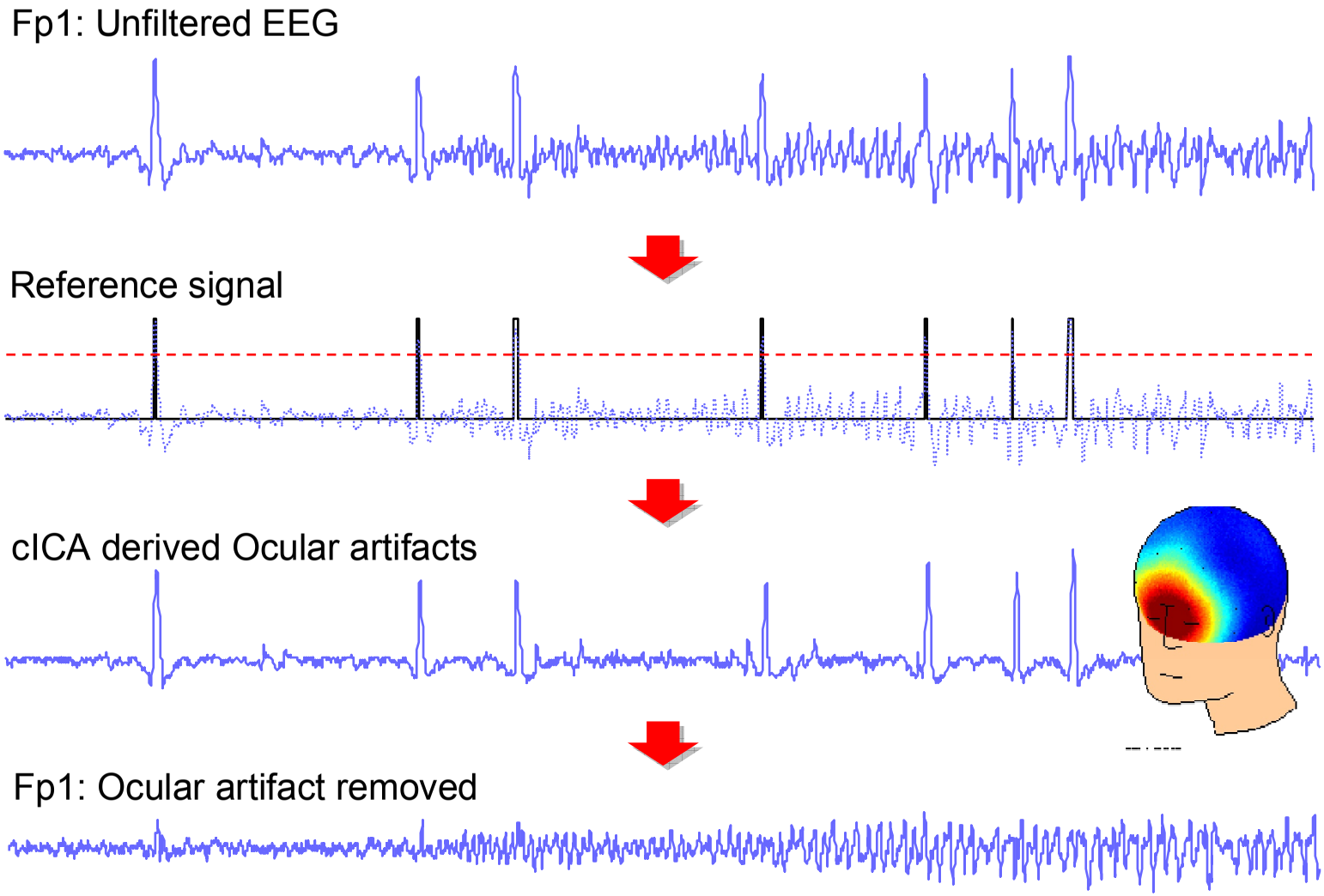
- We have only considered noiseless, square mixing. Other relevant problems which ICA could be used for, include:
  - **Non-square mixing:** n<m
  - **Overcomplete ICA:** *m>n*
  - **Noisy ICA:** (a special case of overcomplete ICA)
  - **Nonlinear ICA:** Difficult!
  - **Constrained ICA:** temporal, spatial or time-frequency constraints

Constrained ICA: Temporal Constraint

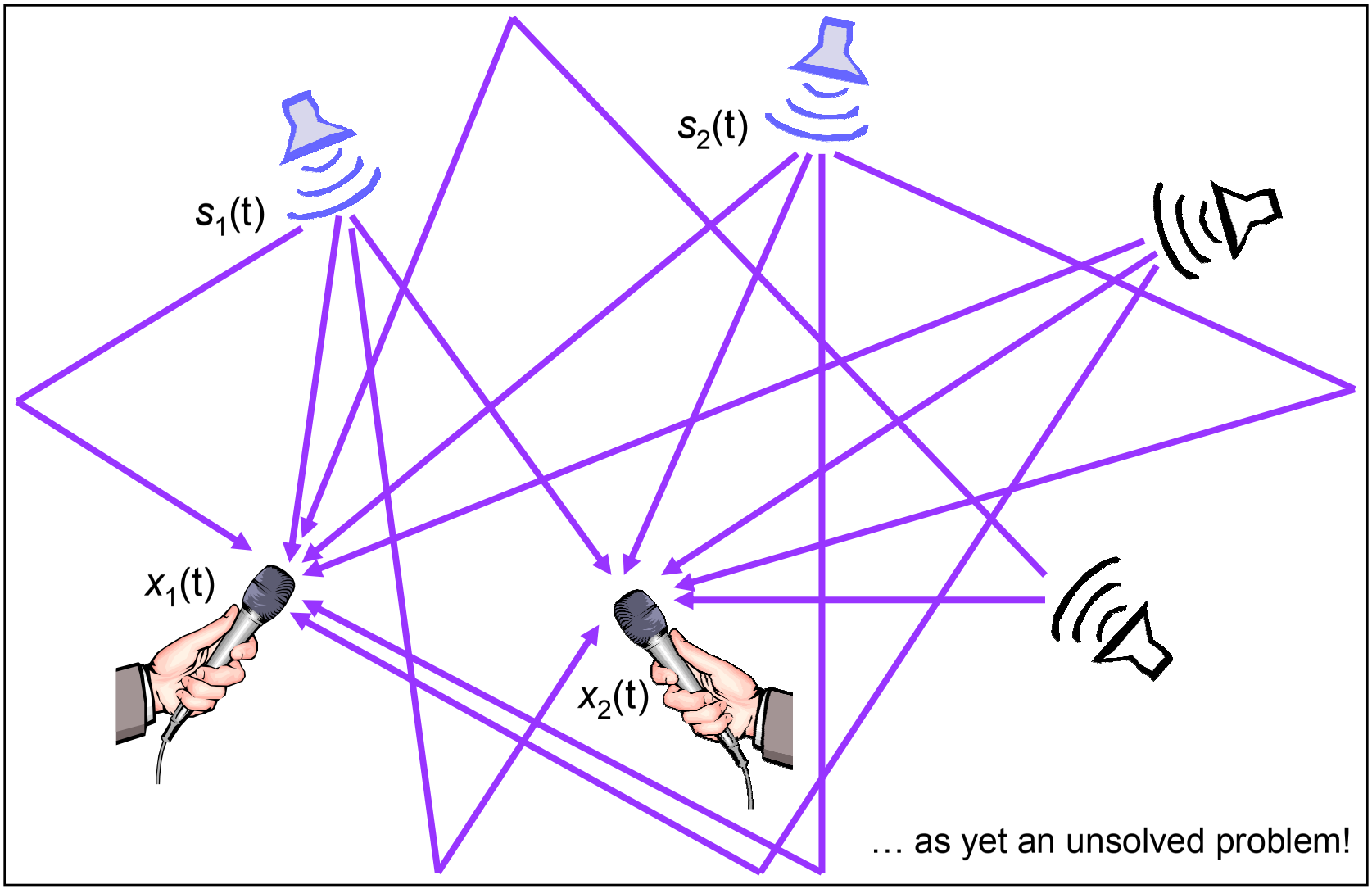


*Based on: Lu and Rajapakse (ICA2001)

Constrained ICA: Removing Eyeblinks



The real ‘Cocktail Party’ problem



Summary

- BSS cannot be achieved using just PCA (i.e. decorrelation)
- BSS can work for non-Gaussian sources by exploiting HOS
- Simple adaptive algorithms exist to perform BSS (ICA)
- Can be extended for convolutive mixing (say)
- Extensions to noisy and over- (and under) complete problems are current problems in research

Bibliography

- A. Hyvarinen, J. Karhunen and E. Oja, Independent Component Analysis, John Wiley & Sons, 2001.
- S. Roberts and R. Everson, Independent Component Analysis: Principles and Practice, Cambridge Press, 2001.
- See also: Proceedings of the ICA conferences (1999, 2000, 2001, 2003, 2004)