UDRC Summer School, Guildford, 27-30 June, 2017

Introduction to Source Separation

Jonathon Chambers[1], and Wenwu Wang[2]

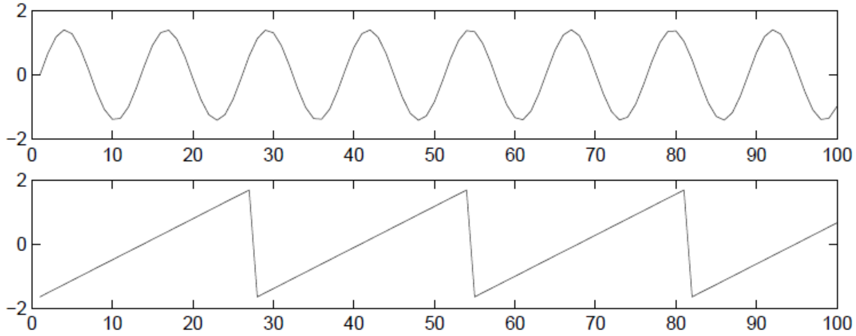
1. School of Electrical and Electronic Engineering, Newcastle University, UK
Jonathon.Chambers@ncl.ac.uk
2. Department of Electrical and Electronic Engineering, University of Surrey, UK
w.wang@surrey.ac.uk

29/06/2017

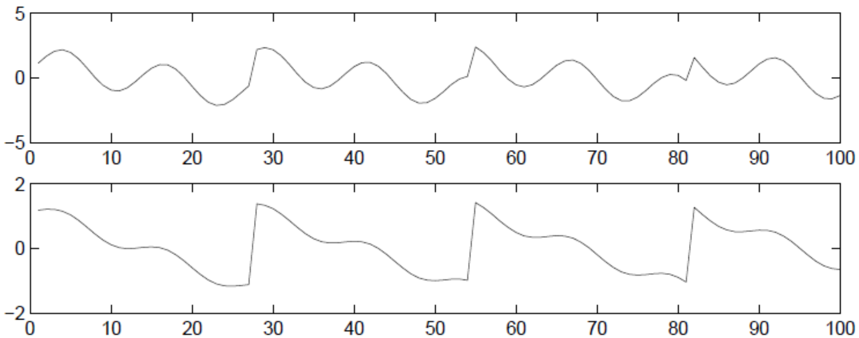
Structure of Talk

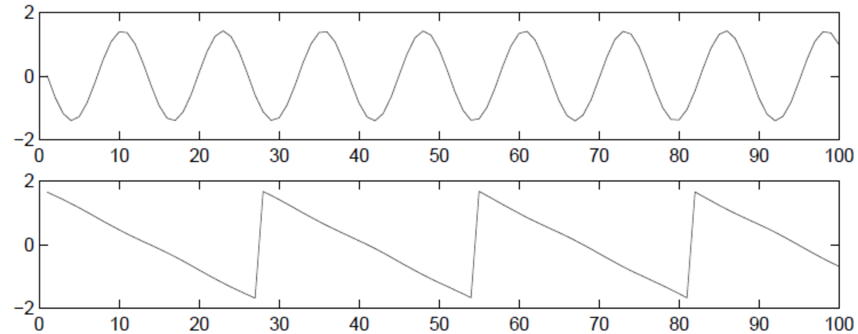
- Introduce the source separation problem and its application domains
- Key books and literature reviews
- Technical preliminaries
- Concepts of ICA – independence and non-Gaussianity
- Types of mixtures
- Taxonomy of algorithms
- Performance measures
- Linear v. non linear unmixing
- Conclusions and acknowledgements

What is Source Separation? - An Example



Two original signals (unknown sources)



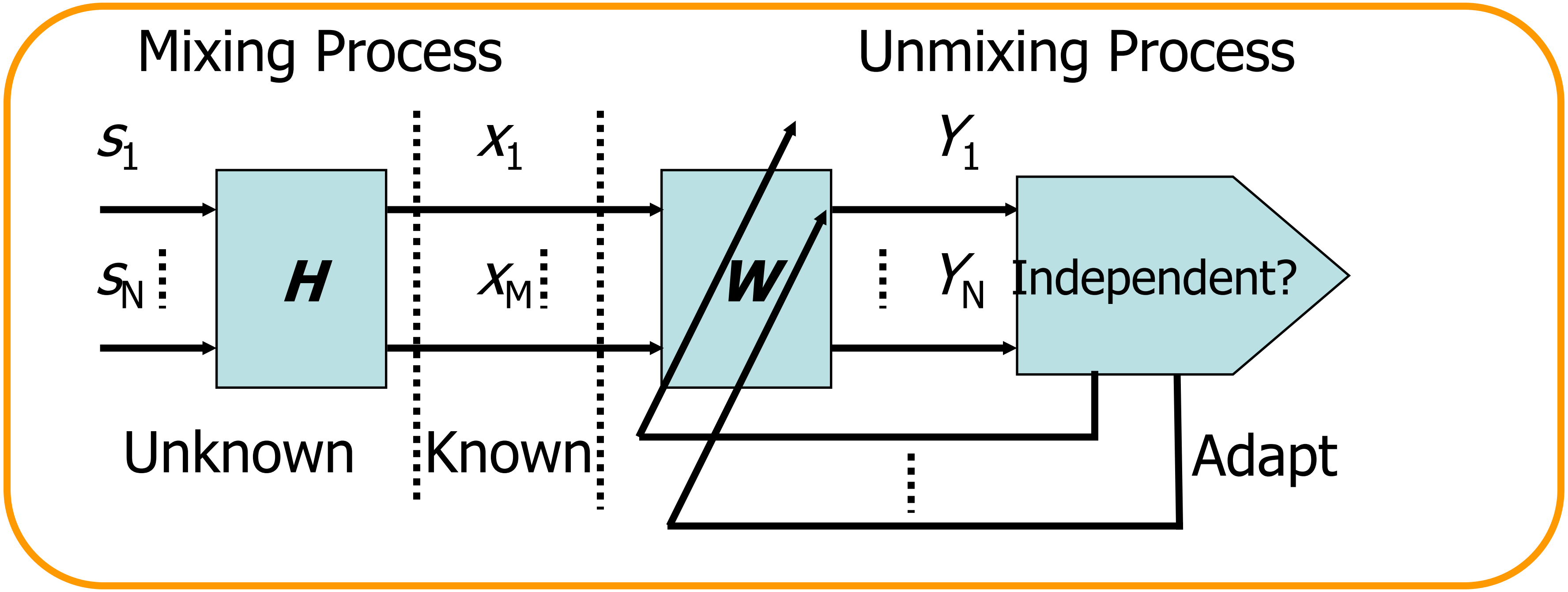


Two observed signals (known mixtures, recorded by sensors)

Estimates of the original source signals

Aapo Hyvarinen and Erkki Oja, Independent component analysis: algorithms and applications, *Neural Networks*, vol. 13, pp. 411-430, 2000

Fundamental Model for ICA/Blind Source Separation

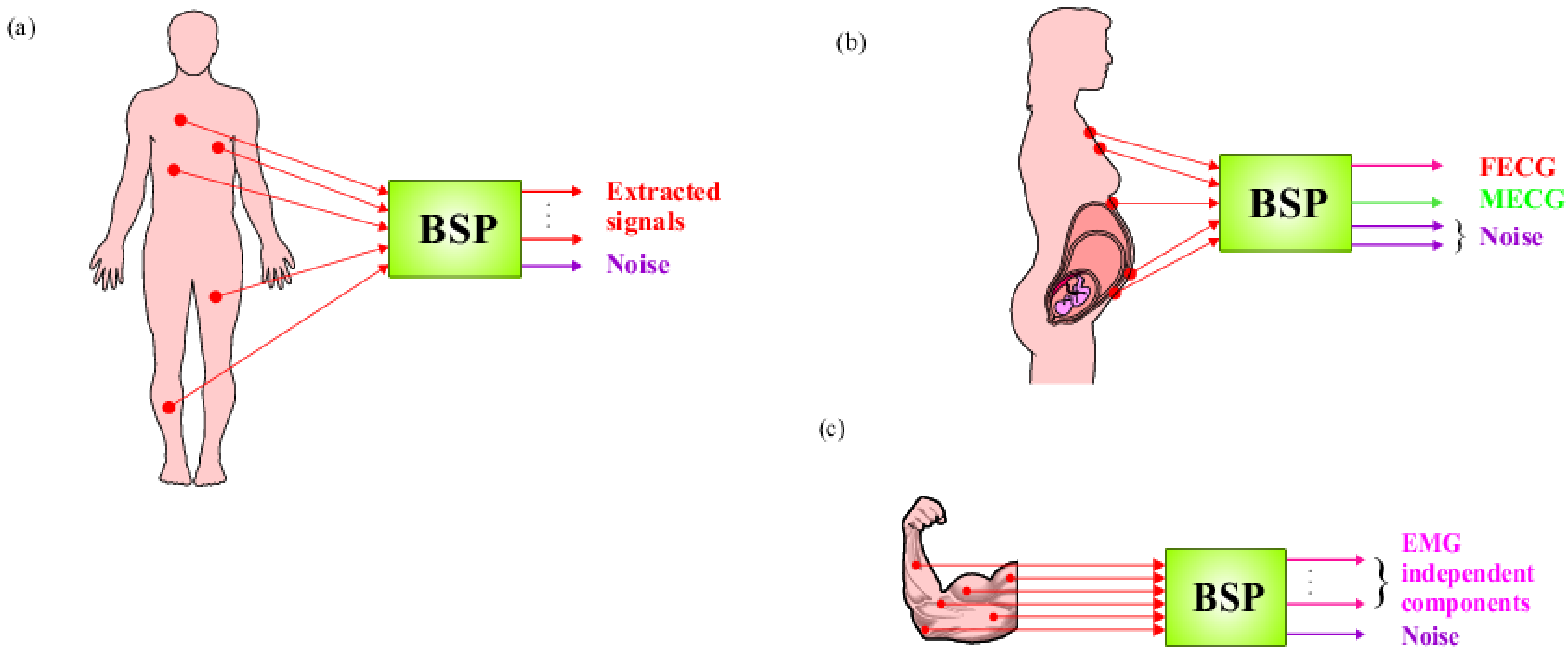


Potential Application Domains

Biomedical signal processing

- Electrocardiography (ECG, FECG, and MECG)
- Electroencephalogram (EEG)
- Electromyography (EMG)
- Magnetoencephalography (MEG)
- Magnetic resonance imaging (MRI)
- Functional MRI (fMRI)

Biomedical Signal Processing



(a) Blind separation for the enhancement of sources, cancellation of noise, elimination of artefacts

(b) Blind separation of FECG and MECG

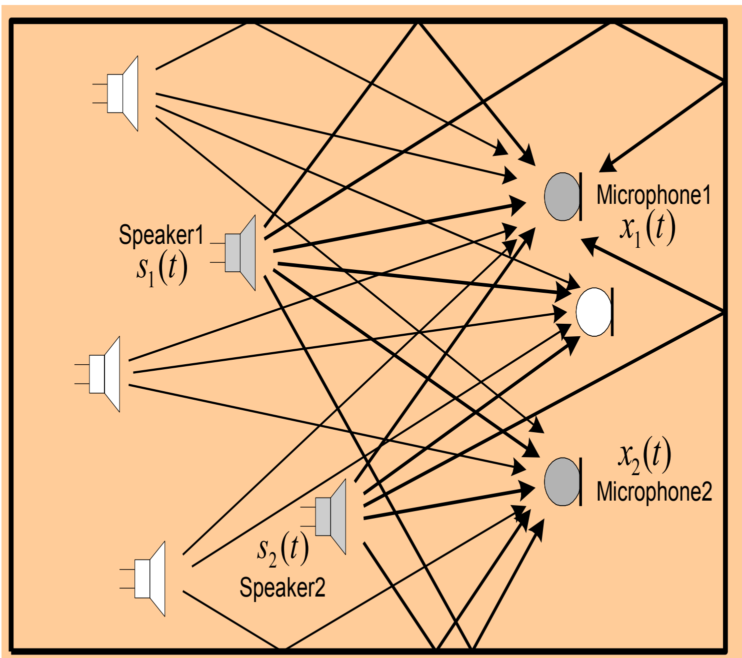
(c) Blind separation of multichannel EMG [Ack. A. Cichocki]

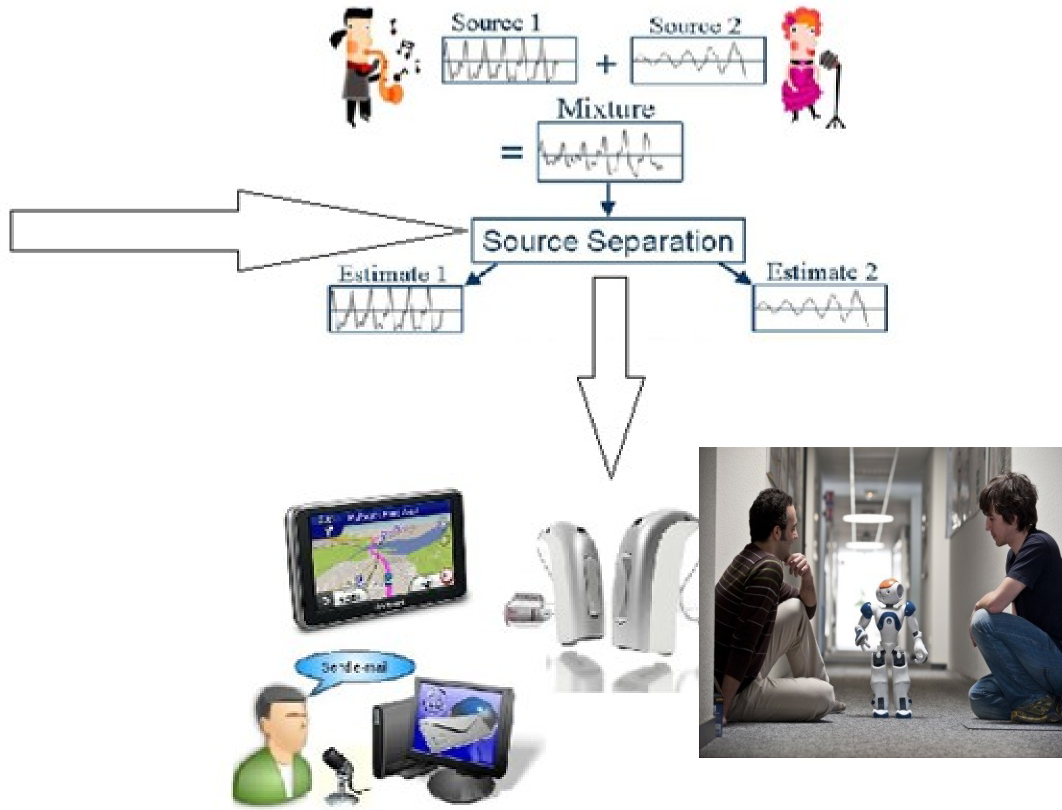
Audio Signal Processing

Cocktail party problem

- Speech enhancement
- Crosstalk cancellation
- Convolutive source separation

Objective of Machine-based Source Separation

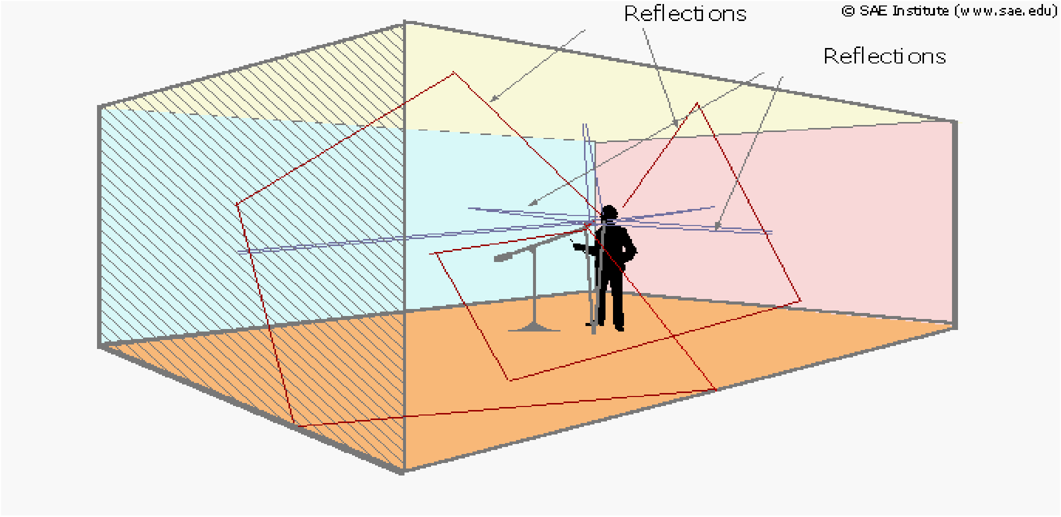


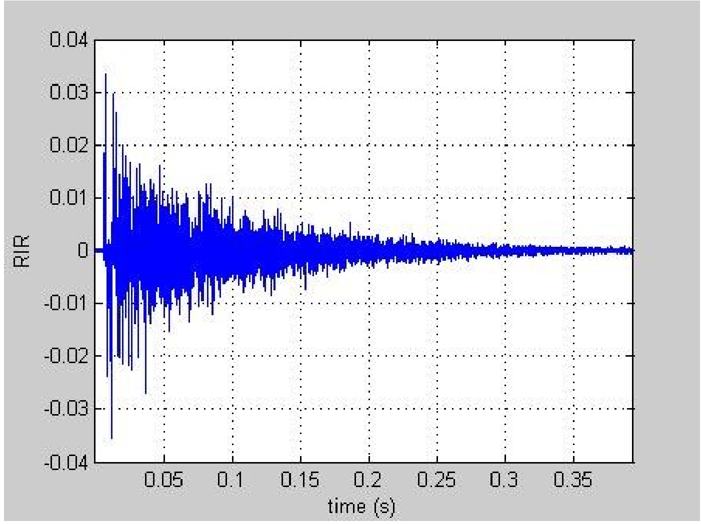


- Scene analysis;
- Hearing aids;
- Robot audition;
- Human computer interaction

The Convolutive Source Separation Problem

- The mixing process is **convolutive**!





A typical room impulse response (RIR)

- **Room reverberation**: multiple reflections of the sound on wall surfaces and objects in an enclosed environment

- **Source separation becomes more challenging as the level of reverberation increases!!**

Communications & Defence Signal Processing

- Multiuser/multi-access communications systems
- Multi-sensor sonar/radar systems
- Digital radio with spatial diversity
- High speed digital subscriber lines

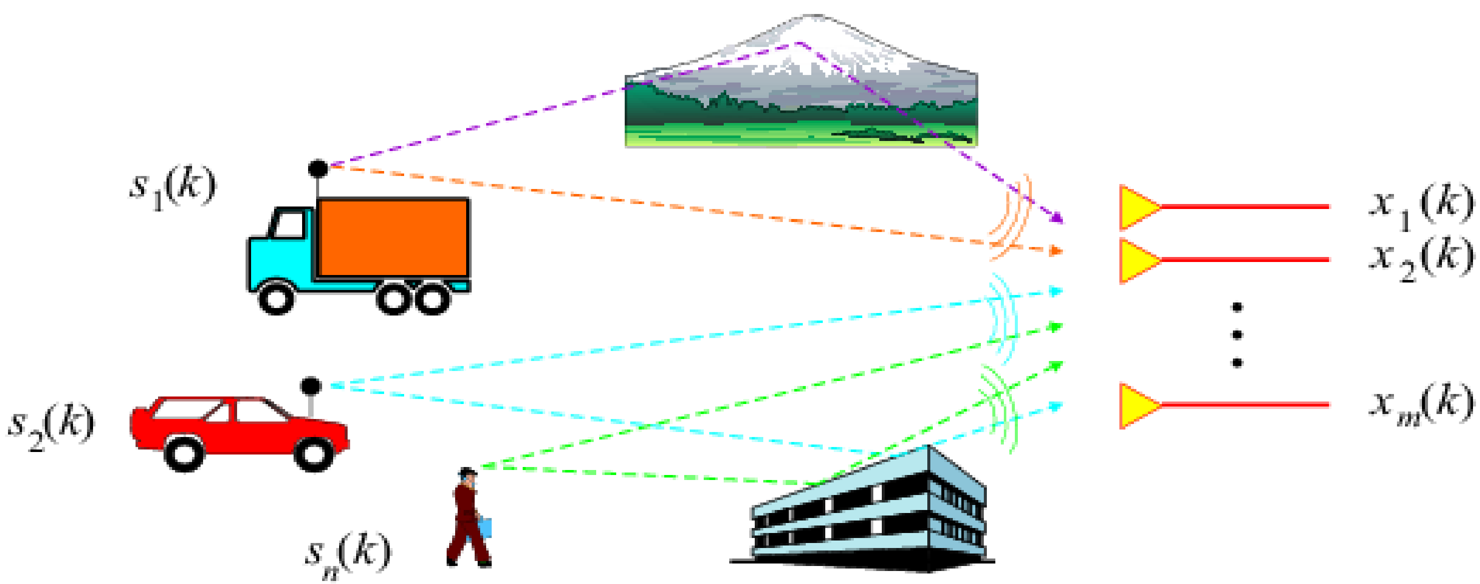
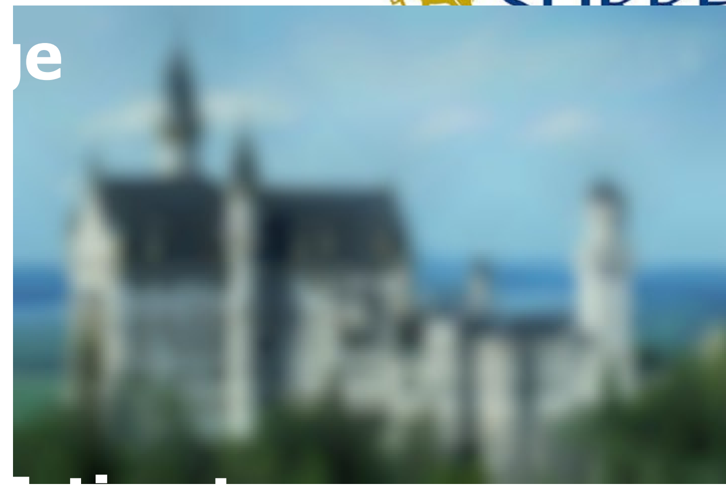


Image Processing

- **Image restoration** (removing blur, clutter, noise, interference etc. from the degraded images)

- **Image understanding** (decomposing the image to basic independent components for scene analysis and recognition)

Blind Image Restoration



0.01	0.01	0.01	0.01	0.01
0.01	0.06	0.10	0.06	0.01
0.01	0.10	0.20	0.10	0.01
0.01	0.06	0.10	0.06	0.01
0.01	0.01	0.01	0.01	0.01

Difference

Key Books and Reviews

- Ganesh Naik and Wenwu Wang, Editors, *Blind Source Separation: Advances in Theory, Algorithms and Applications*, Springer, 2014.
- Pierre Comon and Christian Jutten, Editors, *Handbook of Blind Source Separation Independent Component Analysis and Applications*, New York Academic, 2010.
- Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan and Shun-Ichi Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation,* Wiley 2009.
- Paul D. O'Grady, Barak A. Pearlmutter, and Scott T. Rickard, "Survey of Sparse and Non-Sparse Methods in Source Separation", *Int. Journal of Imaging Systems and Technology*, Vol.15, pp. 20-33, 2005.
- Andrzej Cichocki and Shun-Ichi Amari, *Adaptive Blind Signal and Image Processing*, Wiley, 2002.
- Aapo Hyvärinen, Juha Karhunen and Erkki Oja, *Independent Component Analysis*, Wiley, 2001.

Technical Preliminaries:-
Temporal/Spatial Covariance Matrices
(zero-mean WSS signals)

$$R_x(p) = E\{\underline{x}(t)\underline{x}^T(t-p)\}$$

$$\underline{x}(t) = [x(t)\ x(t-1) \ldots x(t-N+1)]^T$$

$$(\textit{Temporal}\ \text{vector})$$

$$R_{xx} = E\{\underline{x}(t)\underline{x}^T(t)\}$$

$$\underline{x}(t) = [x_1(t)\ x_2(t) \ldots x_N(t)]^T$$

$$(\textit{Spatial}\ \text{vector})$$

Linear Algebra

Linear equation: $$\boxed{\mathbf{Hs} = \mathbf{x}}$$

where:

$$\mathbf{H} = [h_{ij}] \in \Re^{m \times n},$$ known

$$\mathbf{s} \in \Re^{n},$$ unknown

$$\mathbf{x} \in \Re^{m},$$ known

$m=n$, exactly determined

$m>n$, over determined

$m<n$, under determined (or overcomplete)

Linear Equation-: Exactly Determined Case

When $m = n$:

If **H** is non-singular, the solution is uniquely defined by:

$$\boxed{\mathbf{s} = \mathbf{H}^{-1}\mathbf{x}}$$

If **H** is singular, then there may either be no solution (the equations are inconsistent) or many solutions.

Linear Equation :-
Over determined Case

When $m > n$:

If the **H** is full rank (or the columns of **H** are linearly independent), then we have the least squares solution:

$$\boxed{\mathbf{s} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{x}}$$

This solution is obtained by minimization of the norm of the error (exploiting orthogonality principle):

$$\|\mathbf{e}\|^2 = \|\mathbf{x} - \mathbf{H}\mathbf{s}\|^2$$

Linear Equation :- Underdetermined Case

When $m < n$:

There are many vectors that satisfy the equations, and a unique solution is defined to satisfy the minimum norm condition:

$$\min \|\mathbf{s}\|$$

If **H** has full rank, then minimum norm solution is (pseudo inverse):

$$\mathbf{s} = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1} \mathbf{x}$$

Permutation and Scaling Matrices

Permutation matrix:

(an example: 5x5)

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Scaling matrix:

(an example: 5x5)

$$\Lambda = \begin{bmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \lambda_3 & & \\ & & & \lambda_4 & \\ & & & & \lambda_5 \end{bmatrix}$$

Conventional ICA Techniques for Blind Source Separation

H is unknown, i.e. no prior information about **H**

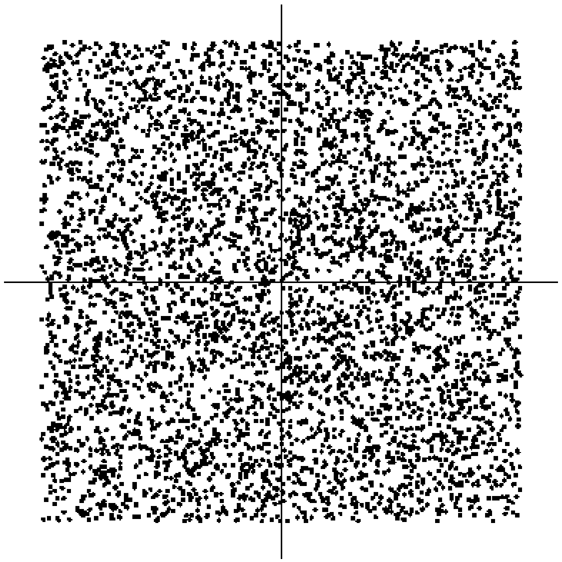
Solution - making assumptions:

1. The sources are *statistically (mutually) independent* from each other.

2. The mixing matrix **H** is a full rank matrix with *m* no less than *n*.

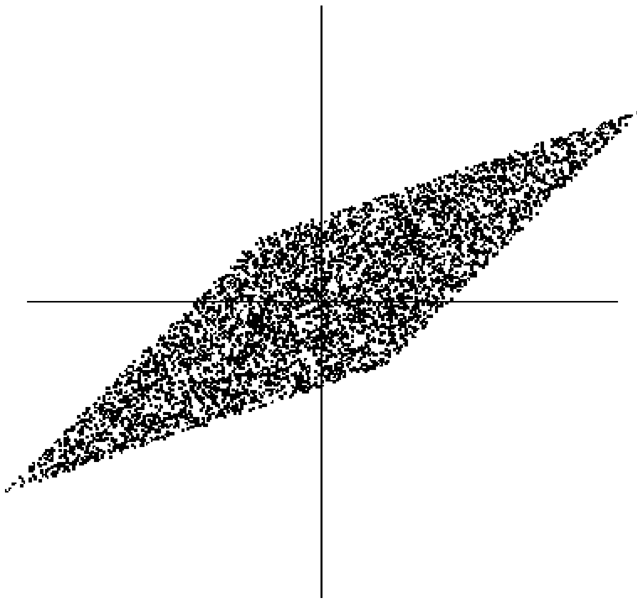
3. At most one source signal has Gaussian distribution.

Illustration of ICA



Joint distribution of two independent components s_1 and s_2 that are uniformly distributed. These two componenents are mixed using a mixing matrix **H** = [2 3; 2 1] to obtain the mixed variables x_1 and x_2.

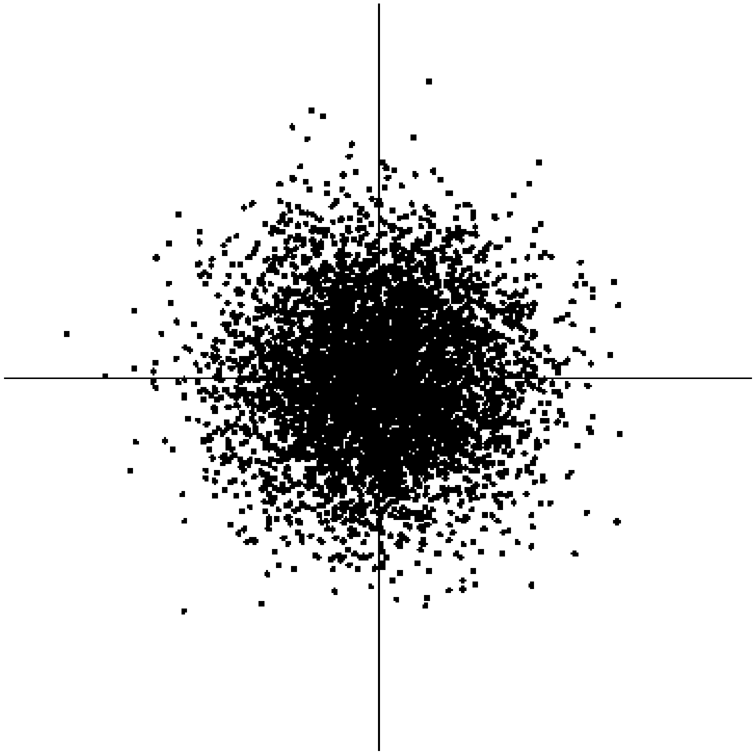
$$p(s_i) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{if } |s_i| \leq \sqrt{3} \\ 0 & \text{otherwise} \end{cases}$$



Joint distribution of the two mixtures x_1 and x_2 which are still uniformly distributed on the parallelogram. By finding the edges, we can potentially estimate the mixing matrix **H**. However, for other distributions this would become much more complicated.

Aapo Hyvarinen and Erkki Oja, Independent component analysis: algorithms and applications, *Neural Networks*, vol. 13, pp. 411-430, 2000

Why non-Gaussianity?



The joint distribution of x_1 and x_2 when the sources s_1 and s_2 are both Gaussian. This figure shows that the joint density is symmetric and does not give any information about the direction of the columns of the mixing matrix **H**.

Aapo Hyvarinen and Erkki Oja, Independent component analysis: algorithms and applications, *Neural Networks*, vol. 13, pp. 411-430, 2000

Maximizing non-Gaussianity Gives Independent Components

- Central Limit Theorem: the distribution of a sum of independent random variables tends toward a Gaussian distribution.
- How could we use the Central Limit Theorem to estimate the mixing matrix **H** then?

$$y = \mathbf{w}^T \mathbf{x} = \mathbf{w}^T \mathbf{H} \mathbf{s} = \mathbf{z}^T \mathbf{s}$$

- From this equation, we can see that y is always more Gaussian than **s**.
- It is clear that if only one of the elements z_i of $\boldsymbol{z}$ is nonzero, we would get the least-Gaussian y.
- In n-dimensional space (i.e. n sources), **w** would have $2n$ local maxima ("2" here comes from the sign ambiguity). To more quickly find these local maxima, a whitening process is often employed to make the subsequent estimate uncorrelated with the previously obtained ones.

Indeterminacies and Ambiguities with the Model

Separation process:

$$\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{H}\mathbf{s} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{s}$$

Separation matrix

Permutation matrix

Scaling matrix

Independence Measurement

Kurtosis (fourth-order cumulant for the measurement of non-Gaussianity):

$$\boxed{kurt(y) = E(y^4) - 3(E(y^2))^2}$$

In practice, we need to find out the direction where the kurtosis of *y* grows most strongly (super-Gaussian signals) or decreases most strongly (sub-Gaussian signals).

(To be covered in detail by Mohsen Naqvi)

Independence Measurement-Cont.

Newcastle University

UNIVERSITY OF SURREY

Mutual information (MI):

$$I(y_1,\cdots,y_n)=\sum_{i=1}^{n}H(y_i)-H(\mathbf{y})\geq 0$$

$$\text{where,}\quad H(\mathbf{y})=\int p(\mathbf{y})\log(p(\mathbf{y}))d\mathbf{y}$$

In practice, minimization of MI leads to the statistical independence between the output signals.

(To be covered in detail by Mohsen Naqvi)

Independence Measurement-Cont.

Kullback-Leibler (KL) divergence:

$$KL\left[p(\mathbf{y}) \,\|\, \prod \left(p_{y_i}(y_i)\right)\right] = \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod \left(p_{y_i}(y_i)\right)} d\mathbf{y}$$

Minimization of KL between the joint density and the product of the marginal densities of the outputs leads to the statistical independence between the output signals.

Types of Sources

- Non-Gaussian signals (super/sub-Gaussian) [Conventional BSS]
- Stationary signals [Conventional BSS]
- Temporally correlated but spectrally disjoint signals [SOBI, Belouchrani et al., 1993]
- Non-stationary signals [Freq. Domain BSS, Parra & Spence, 2000]
- Sparse Signals [Mendal, 2010]

Types of Mixtures

- Instantaneous mixtures (memory-less, flat fading):

$$\mathbf{x} = \mathbf{H}\mathbf{s}$$ (Direct form)

A scalar matrix

$$\mathbf{x} = \mathbf{s}^T \mathbf{H}^T$$ (Transpose form)

- Convolutive mixtures (with indirect response with time-delays)

$$\mathbf{x} = \mathbf{H} * \mathbf{s}$$

A filter matrix

Types of Mixtures-Cont.

- Noisy and non negative mixtures (corrupted by noises and interferences):

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \underbrace{\mathbf{n}}_{\text{Noise vector}}$$

$$\text{where } \mathbf{H} \geq \mathbf{0} \text{ and } \mathbf{s} \geq \mathbf{0}$$

- Non-linear mixtures (mixed with a mapping function)

$$\mathbf{x} = \underbrace{F}_{\text{Unknown nonlinear function}}(\mathbf{s})$$

Taxonomy of Algos. :- Block Based- JADE

Joint Approximate Diagonalization of Eigen-matrices (JADE) (Cardoso & Souloumiac, 1993):

1. Initialisation. Estimate a whitening matrix **V**, and set $\bar{\mathbf{x}} = \mathbf{V}\mathbf{x}$

2. Form statistics. Est. set of 4th order cumulant matrices: $\mathbf{Q}_i$

3. Optimize an orthogonal contrast. Find the rotation matrix **U** such that the cumulant matrices are as diagonal as possible (using Jacobi rotations), that is

$$\mathbf{U} = \arg\min_{\mathbf{U}} \left(\mathit{off}\left(\sum_i \mathbf{U}^H \mathbf{Q}_i \mathbf{U}\right) \right)$$

4. The separation matrix is therefore obtained by rotation & whitening:

$$\mathbf{W} = \mathbf{U}^{-1}\mathbf{V} = \mathbf{U}^H\mathbf{V}$$

Taxonomy of algorithms:- Block Based - SOBI.

Second Order Blind Identification (SOBI) (Belouchrani et al., 1993):

1. Perform robust orthogonalization: $\bar{\mathbf{x}}(k) = \mathbf{V}\mathbf{x}(k)$

2. Estimate the set of covariance matrices:

$$\hat{\mathbf{R}}_{\bar{\mathbf{x}}}(p_i) = (1/N)\sum_{k=1}^{N} \bar{\mathbf{x}}(k)\bar{\mathbf{x}}^T(k-p_i) = \mathbf{V}\hat{\mathbf{R}}_{\mathbf{x}}(p_i)\mathbf{V}^T$$

where p_i is a pre-selected set of time lags.

3. Perform joint approximate diagonalization:

$$\hat{\mathbf{R}}_{\bar{\mathbf{x}}}(p_i) = \mathbf{U}\mathbf{D}_i\mathbf{U}^T$$

4. Estimate the source signals:

$$\hat{\mathbf{s}}(k) = \mathbf{U}^T\mathbf{V}\mathbf{x}(k)$$

Taxonomy of Algorithms:- Block Based - FastICA

Fast ICA (Hyvärinen & Oja, 1999):

1. Choose an initial (e.g. random) weighting vector **W**

2. Let $\mathbf{W}^{+} = E\left\{\mathbf{x}g\left(\mathbf{W}^{T}\mathbf{x}\right)\right\} - E\left\{\dot{g}\left(\mathbf{W}^{T}\mathbf{x}\right)\right\}\mathbf{W}$

   Non-linearity g(.) chosen as a function of sources.

3. Let $\mathbf{W} = \mathbf{W}^{+} / \left\|\mathbf{W}^{+}\right\|$

4. If not converged, go back to step 2.

(Details to be covered by Mohsen Naqvi)

Taxonomy of Algos:- Sequential - InforMax

InforMax (Minimal Mutual Information/Maximum Entropy) (Bell & Sejnowski, 1995):

$$J_{MMI}(\mathbf{W}) = \sum_i h_i(y_i, \mathbf{W}) - h(\mathbf{y}, \mathbf{W})$$

$$= -h(\mathbf{x}) - \log\left|\det(\mathbf{W})\right| - E\left[\sum_i p_{y_i}(y_i, \mathbf{W})\right]$$

$$J_{ME}(\mathbf{W}) = h(\mathbf{z}, \mathbf{W}) = -E[\log p_z(\mathbf{z})] = -E[\log p_z(g(\mathbf{Wx}))]$$

$$= h(\mathbf{x}) + \log\left|\det(\mathbf{W})\right| + \sum_i E[\log(\dot{g}_i(y_i))]$$

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta\left[\mathbf{I} - \varphi(\mathbf{y})\mathbf{y}(k)^T\right]\mathbf{W}(k)$$

Taxonomy of Algos:- Sequential - Natural Gradient

Natural Gradient (Amari & Cichocki, 1998):

In *Riemannian* geometry, the distance metric is defined as:

$$d_w(\mathbf{W}, \mathbf{W} + \delta\mathbf{W}) = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} \delta w_i \delta w_j g_{ij}(\mathbf{W})} = \sqrt{\delta\mathbf{W}^T G(\mathbf{W})\delta\mathbf{W}}$$

General adaptation equation:

$$\mathbf{W}(k+1) = \mathbf{W}(k) - \mu(k) G^{-1}(\mathbf{W}(k)) \frac{\partial J(\mathbf{W}(k))}{\partial \mathbf{W}}$$

Specifically: $$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta\left[\mathbf{I} - f(\mathbf{y})\mathbf{y}^T(k)\right]\mathbf{W}(k)$$

(Details to be covered by Mohsen Naqvi)

Performance Measurement

Performance index (Global rejection index):

$$PI(\mathbf{G}) = \sum_{i=1}^{m}\left(\sum_{j=1}^{m}\frac{|g_{ij}|}{\max_k |g_{ik}|} - 1\right) + \sum_{i=1}^{m}\left(\sum_{j=1}^{m}\frac{|g_{ij}|}{\max_k |g_{kj}|} - 1\right)$$

Waveform matching (Mean Squared Error):

$$\varepsilon^2 = E\left\{\left\|\hat{\mathbf{s}} - \mathbf{s}\right\|^2\right\}$$

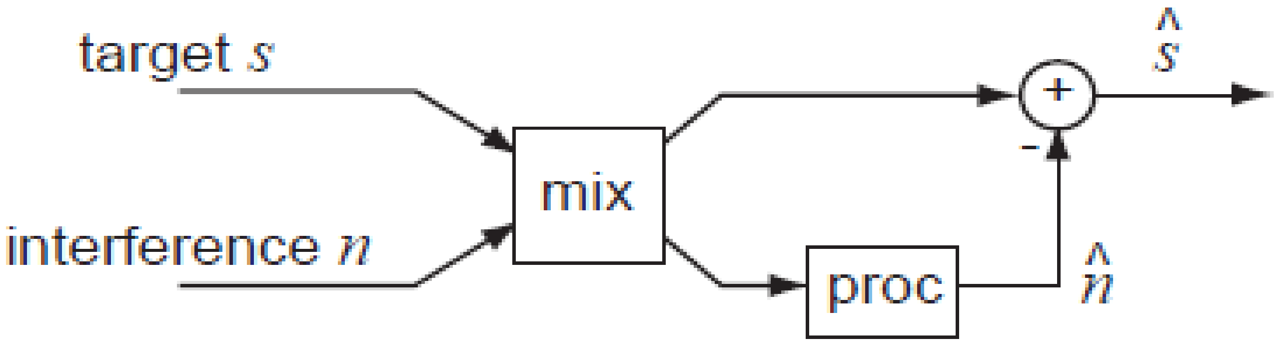
Performance Measurement – Cont.

- Signal to Interference Ratio
- Signal to Artefact Ratio
- Signal to Distortion Ratio

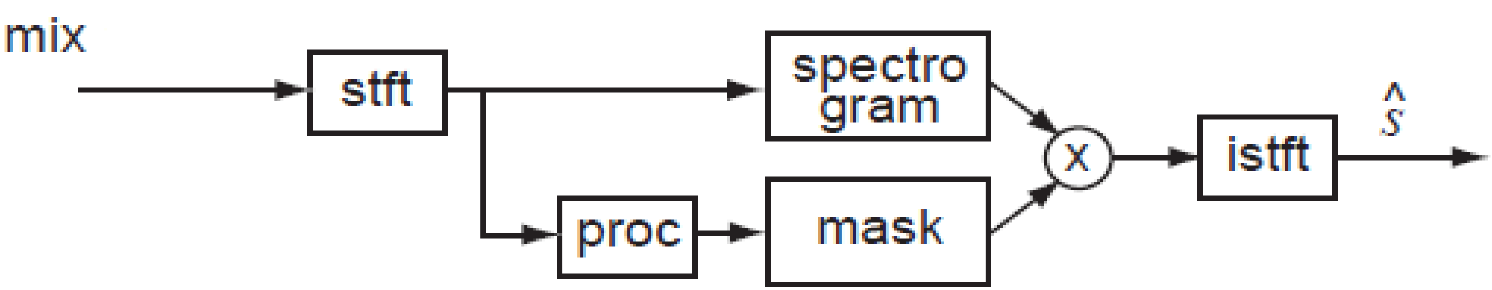
- Perceptual Evaluation Speech Quality (PESQ)
- Perceptual Evaluation Audio Quality (PEAQ)
- Perceptual Evaluation of Audio Source Separation (PEASS)

From Time to Time-Frequency Domain

- Time domain: Multichannel ICA/Beamforming (more to be discussed by Mohsen Naqvi and Stephan Weiss)

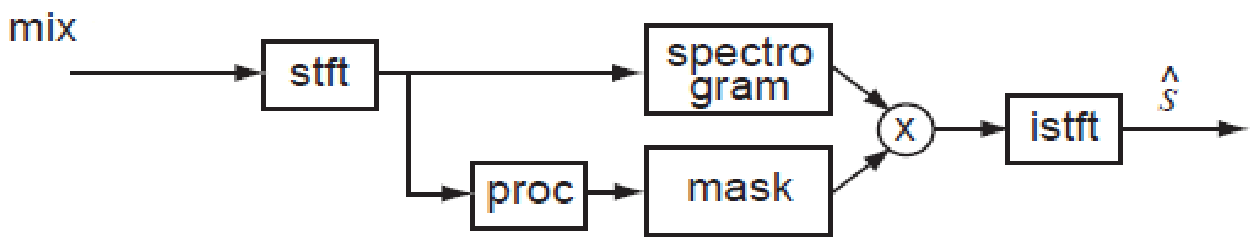


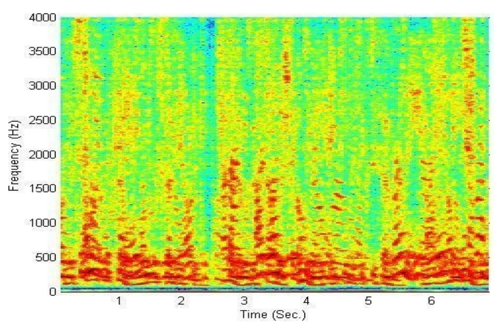
- Time-frequency domain: frequency domain ICA/time frequency masking (more to be covered in my second lecture)



Time-Frequency Masking

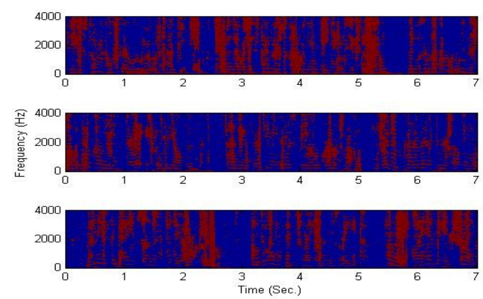
Audio sources can be extracted by simple masking operations



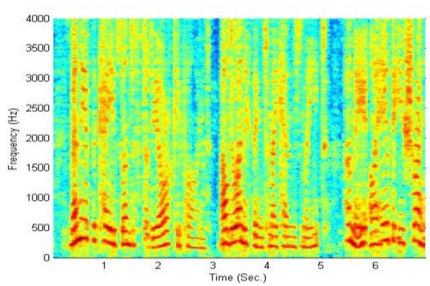


mixture

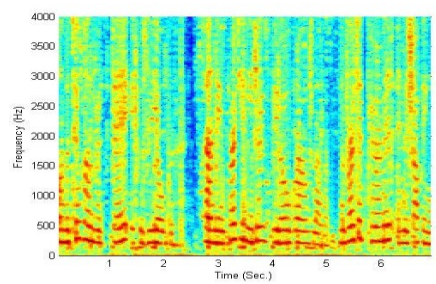
X



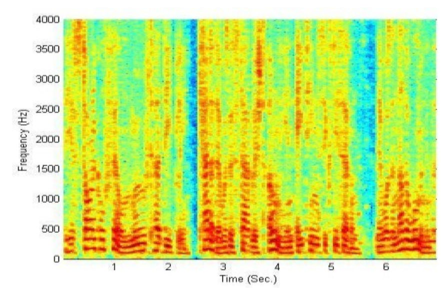
Masks



Source 1



Source 2



Source 3

Other Methods and Recent Trends

- Polynomial matrix decomposition (to be covered by Stephan Weiss)
- Non-negative matrix factorization
- Sparse representations
- Low-rank representation
- Deep neural networks
- Informed/assisted/supervised/semi-supervised source separation
- Interactive (on the fly) source separation
- …

Summary

In this talk, we have reviewed:

- BSS applications and concepts
- Mathematical preliminaries
- Type of sources and mixtures
- Representative block and sequential algorithms
- Performance measures
- Transform domain separation
- Other methods and recent trends

Some of these will be discussed in more depth in the ensuing talks.

Acknowledgements

We wish to express our sincere thanks for the support of Professor Andrzej Cichocki, Riken Brain Science Institute, Japan, and cites the use of some of the figures in his book in this talk.

The invitation to give this part of the vacation school.

References

- J.-F. Cardoso and A. Souloumiac. "Blind beamforming for non Gaussian signals", In *IEE Proceedings-F*, vol. 140, no. 6, pp. 362-370, December 1993.
- A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, E. Moulines. "A blind source separation technique based on second order statistics", *IEEE Trans. on Signal Processing,* vol. 45, no 2, pp. 434-44, Feb. 1997.
- A. Mansour and M. Kawamoto, "ICA Papers Classified According to their Applications and Performances", *IEICE Trans. Fundamentals*, vol. E86-A, no. 3, pp. 620-633, March 2003.
- Aapo Hyvärinen, "Survey on Independent Component Analysis", *Neural Computing Surveys*, vol. 2, pp. 94-128, 1999.
- A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis", *Neural Computation*, vol. 9, no. 7, pp. 1483-1492, 1997.
- J. Bell, and T. J. Sejnowski, "An information-maximization to blind separation and blind deconvolution", *Neural Comput.,* vol. 7, pp. 1129-1159, 1995.
- S. Amari, A. Cichocki, and H.H. Yang, "A new learning algorithm for blind source separation". In *Advances in Neural Information Processing 8*, pp. 757-763. MIT Press, Cambridge, MA, 1996.
- L. Parra, and C. Spence, "Convolutive blind separation of non-stationary sources", *IEEE Trans. Speech Audio Processing*, vol. 8, no. 3, pp. 320–327, 2000.
- T.-W. Lee, *Independent Component Analysis: Theory and Applications*, Kluwer, 1998 .
- A. Hyvarinen and E. Oja, Independent component analysis: algorithms and applications, *Neural Networks*, vol. 13, pp. 411-430, 2000.

References

- H. Buchner, R. Aichner, and W. Kellermann, “Blind source separation for convolutive mixtures: A unified treatment”. In Huang, Y. and Benesty, J., editors, *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, pp. 255–293. Kluwer Academic Publishers, 2004.
- S. Araki, S. Makino, A. Mukai Blin, and H. Sawada, “Underdetermined blind separation for speech in real environments with sparseness and ICA”. In *Proc. ICASSP*, volume III, pp. 881–884, 2004.
- M. I. Mandel, S. Bressler, B. Shinn-Cunningham, and D. P. W. Ellis, “Evaluating source separation algorithms with reverberant speech,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1872–1883, 2010.
- Y. Hu and P.C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol.16, no.1, pp.229-238, Jan. 2008.
- Y. Luo, W. Wang, J. A. Chambers, S. Lambotharan, and I. Prouder, "Exploitation of source non-stationarity for underdetermined blind source separation with advanced clustering techniques," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2198-2212, June 2006.
- W. Wang, S. Sanei, and J.A. Chambers, "Penalty function based joint diagonalization approach for convolutive blind separation of nonstationary sources," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1654-1669, May 2005.
- T. Xu, W. Wang, and W. Dai, "Sparse coding with adaptive dictionary learning for underdetermined blind speech separation", *Speech Communication*, vol. 55, no. 3, pp. 432-450, 2013.
- Y. Yu, W. Wang, and P. Han, "Localization based stereo speech source separation using probabilistic time-frequency masking and deep neural networks", *EURASIP Journal on Audio Speech and Music Processing*, 2016:7, 18 pages, DOI 10.1186/s13636-016-0085-x, 2016.