Loughborough Surrey
Strathclyde and Cardiff
(LSSC) Consortium

Fundamentals of ICA and Blind Source Separation

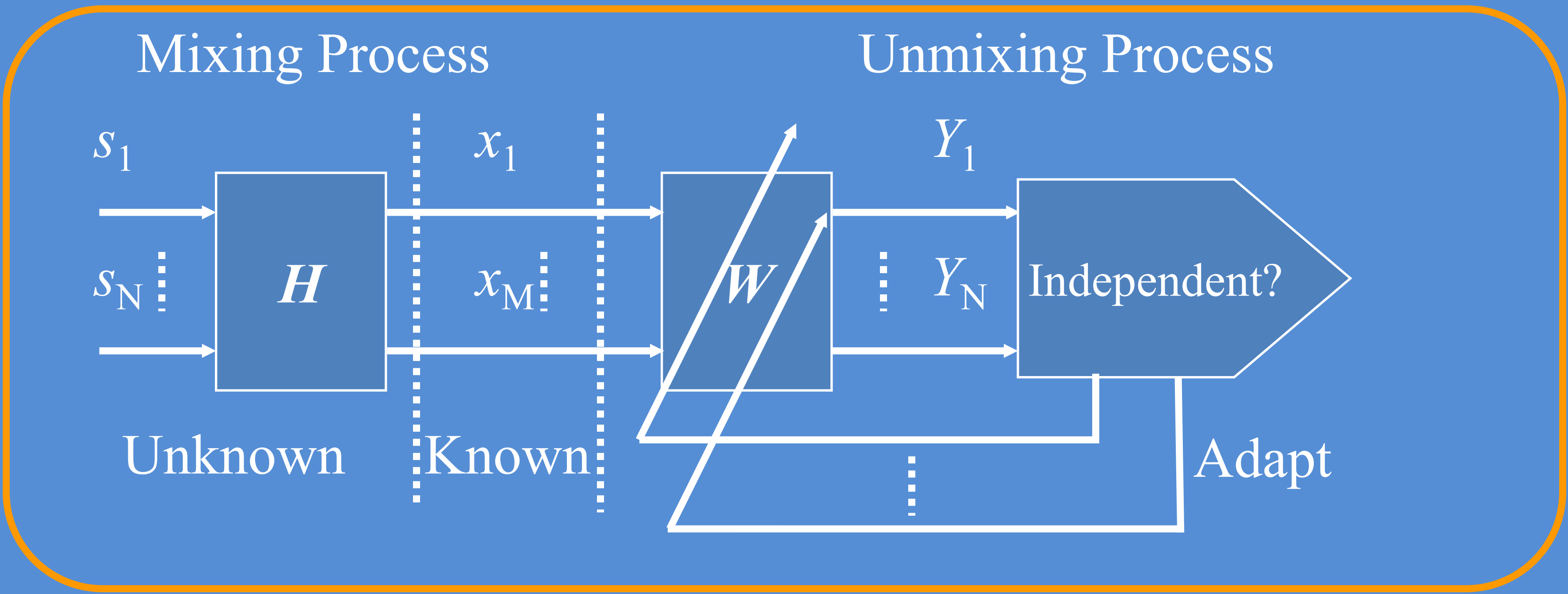
Jonathon Chambers FREng
Head Advanced Signal Processing Group
School of Electronic, Electrical and Systems Engineering
Loughborough University
E-mail: j.a.chambers@lboro.ac.uk

UDRC Vacation School, Heriot-Watt University,
26th June 2014

Structure of Talk

- Introduce the source separation problem and its application domains
- Key books and literature reviews
- Technical preliminaries
- Type of mixtures
- Taxonomy of algorithms
- Performance measures
- Conclusion
- Acknowledgement

Fundamental Model for Instantaneous ICA/Blind Source Separation

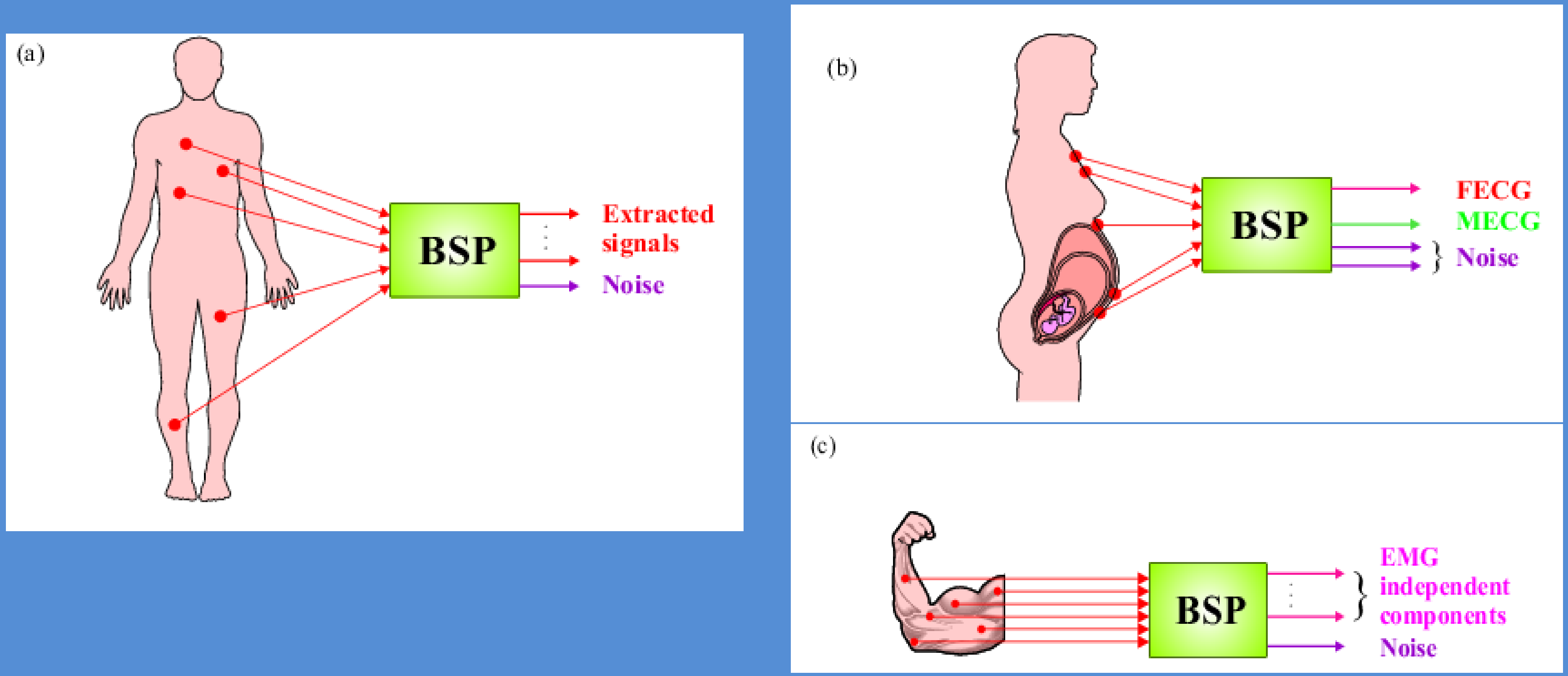


Potential Application Domains

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 artifacts

(b) Blind separation of FECG and MECG

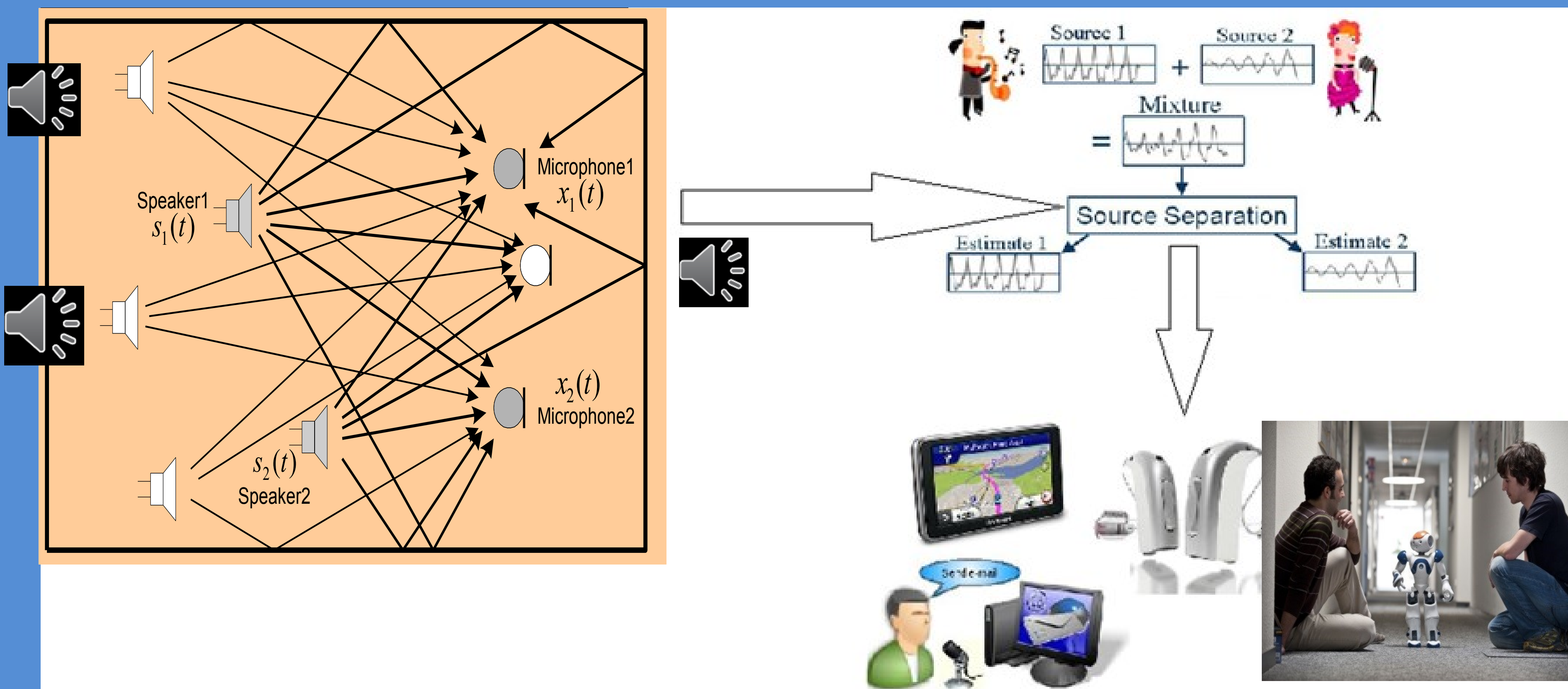
(c) Blind separation of multichannel EMG [Ack. A. Cichocki]

Audio Signal Processing

Cocktail party problem

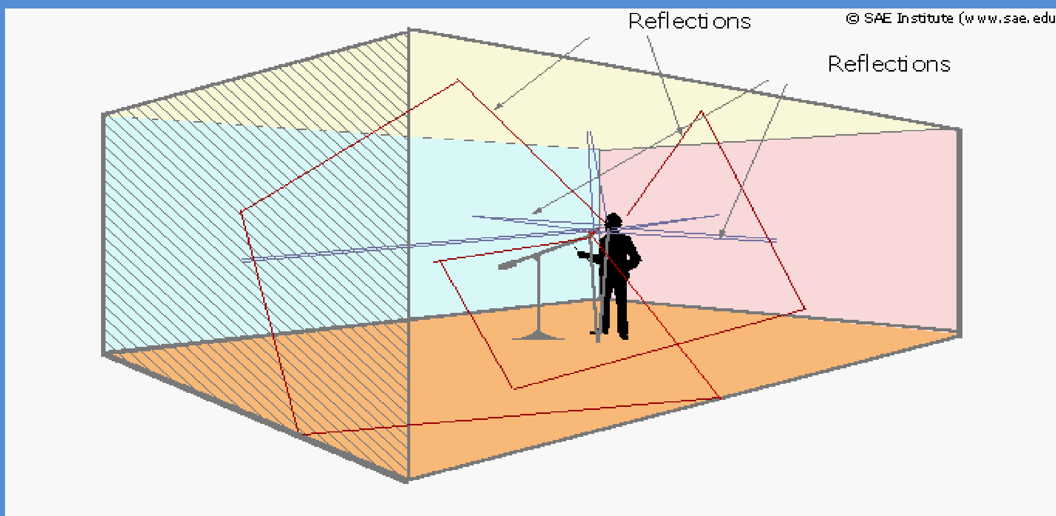
- Speech enhancement
- Crosstalk cancellation
- Convolutive source separation

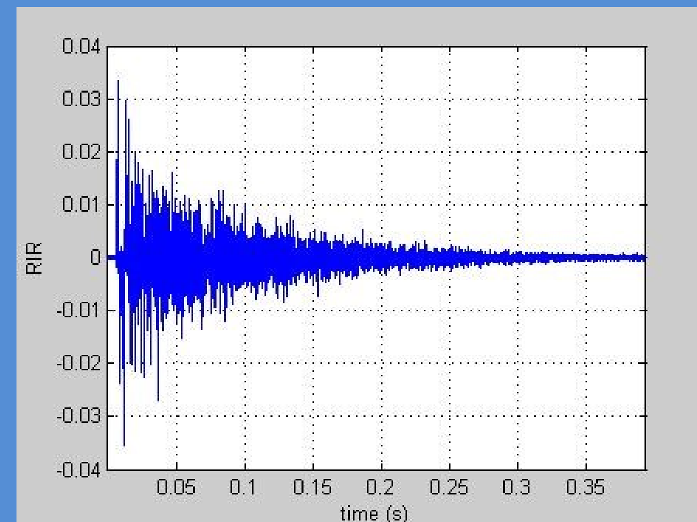
Objective of Machine-based Source Separation



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 enclosure

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

Communications & Defence Signal Processing

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

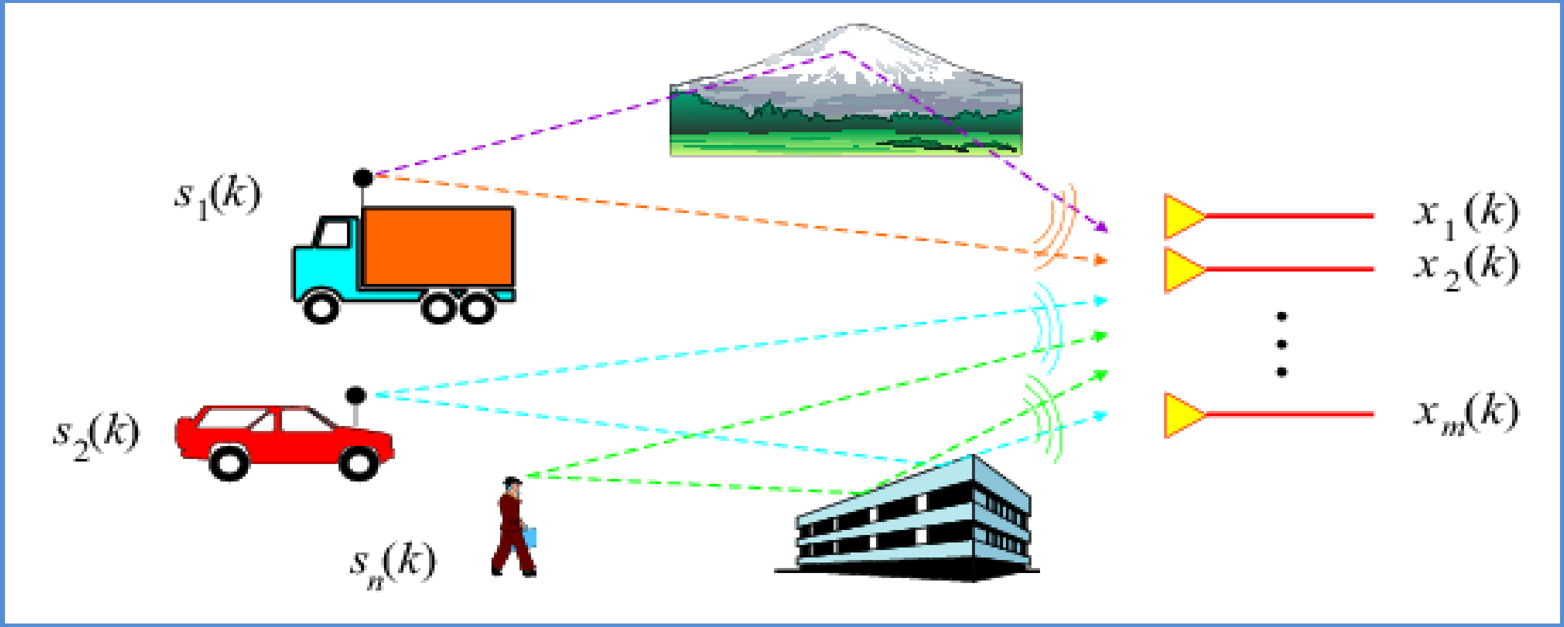


Image Processing

- **Image restoration** (removing blur, clutter, noise, interference etc. from the degraded images)
- **Image understanding** (decomposing the image to basic independent components for sparse representation of image with application to, for example, image coding)

Blind Image Restoration

Degraded Image

Image Estimate

Blur Estimate

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

Temporal/Spatial Covariance Matrices

(zero-mean WSS signals)

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

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

(Temporal vector)

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

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

(Spatial vector)

Technical Preliminaries:- Linear Algebra

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

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 (overcomplete)

Linear Equation-: Exactly Determined Case

When m=n:

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

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

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:

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

This solution is obtained by minimization of the norm of the error (exploit 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):

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

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 Blind Source Separation

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

Solution - making assumptions:

1. The sources are *statistically (mutually) independent* of 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.

Indeterminacies

Separation process:

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

Separation matrix

Permutation matrix

Scaling matrix

Independence Measurement

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

$$kurt(y) = E(y^4) - 3(E(y^2))^2$$

In practice, find out the direction where the kurtosis of y grows most strongly (super-Gaussian signals/Leptokurtic) or decreases most strongly (sub-Gaussian signals/Platykurtic).

Independence Measurement-Cont.

Mutual information (MI):

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

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

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

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 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, Cardoso, 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):

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}$$

Noise vector

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

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

$$\mathbf{x} = F(\mathbf{s})$$

Unknown function

Taxonomy of Algos. :- Block Based- JADE

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

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 rots), that is

$$\mathbf{U} = \arg\min_{\mathbf{U}} \left(\mathrm{off}\left(\sum_{\mathbf{i}} \mathbf{U}^{\mathbf{H}} \mathbf{Q}_{\mathbf{i}} \mathbf{U}\right) \right)$$

4. The separation matrix is therefore obtained unitary (rotation) & whiten.:

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

Taxonomy of Algorithms:- Block Based - SOBI.

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

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 lag

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):

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

$$\mathbf{W} = \mathbf{W}^{+} \Big/ \left\|\mathbf{W}^{+}\right\|$$

4. If not converged, go to step 2.

Taxonomy of Algos:- Sequential - InforMax

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

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

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

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

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

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

Taxonomy of Algos:- Sequential - Natural Gradient

Natural Gradient (Amari & Cichocki):

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

Performance Measurement

Performance index (Global rejection index):

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

Waveform matching:

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

Performance Measure

BSS Eval Toolbox [http://bass-db.gforge.inria.fr/bss_eval/]:

This MATLAB toolbox give reliable results in the form of Source to Interference Ratio (SIR), Source to Distortion Ratio (SDR), Source to Noise Ratio(SNR), and Source to Artifact Ratio (SAR).

$$\mathrm{SIR} = 10log_{10}\frac{||\ \mathrm{S_{target}}\ ||^2}{||\ \mathrm{einterf}\ ||^2}$$

$$\mathrm{SDR} = 10log_{10}\frac{||\ \mathrm{Starget}\ ||^2}{||\ \mathrm{einterf} + \mathrm{enoise} + \mathrm{eartif}\ ||^2}$$

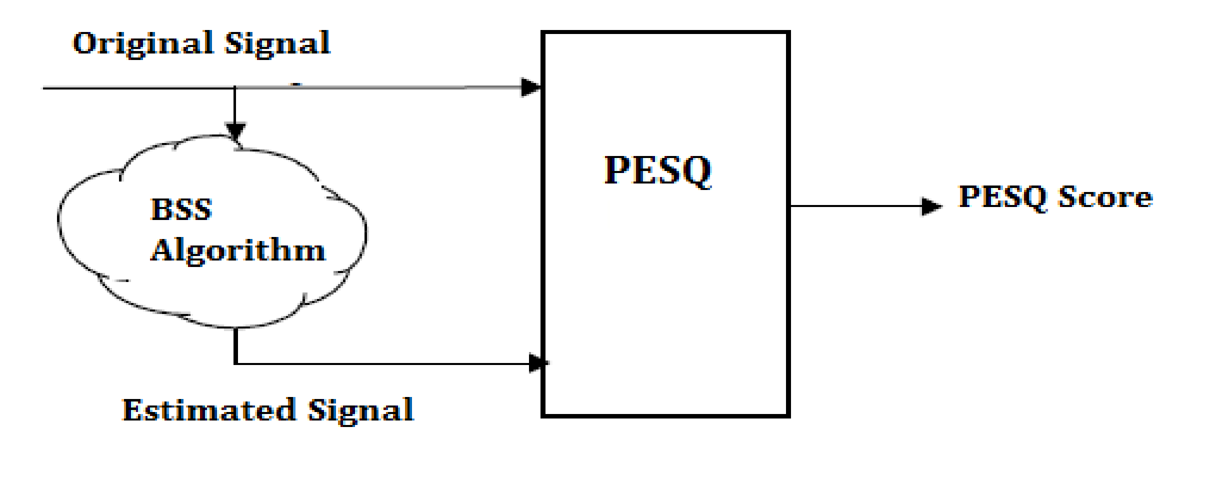
$$\mathrm{SDR} = 10log_{10}\frac{||\ \mathrm{e_{interf}} + \mathrm{enoise} + \mathrm{eartif}\ ||^2}{||\ \mathrm{einterf}\ ||^2}$$

Performance Measure

Perceptual Evaluation Speech Quality:

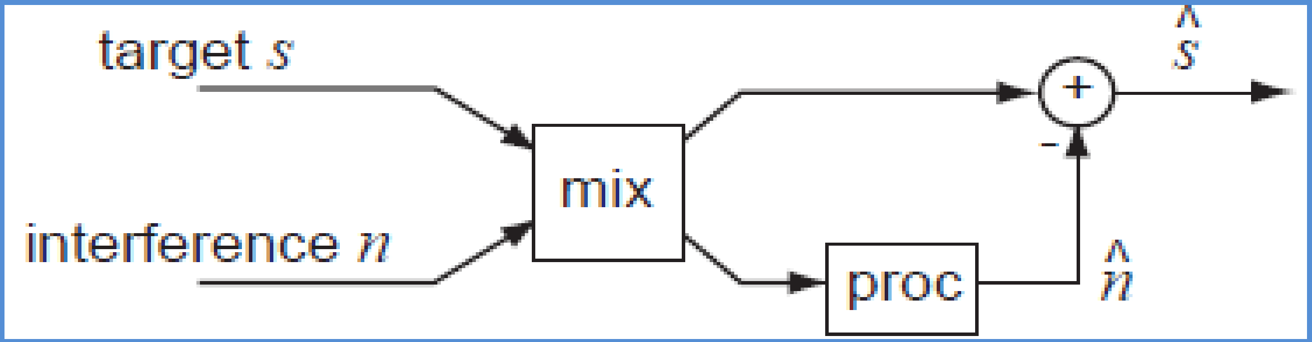
This is basically an algorithm that is design to predict subjective opinion scores of a degraded audio sample.

It give us the Mean Opinion Score for the speech quality, that values from 0-5.

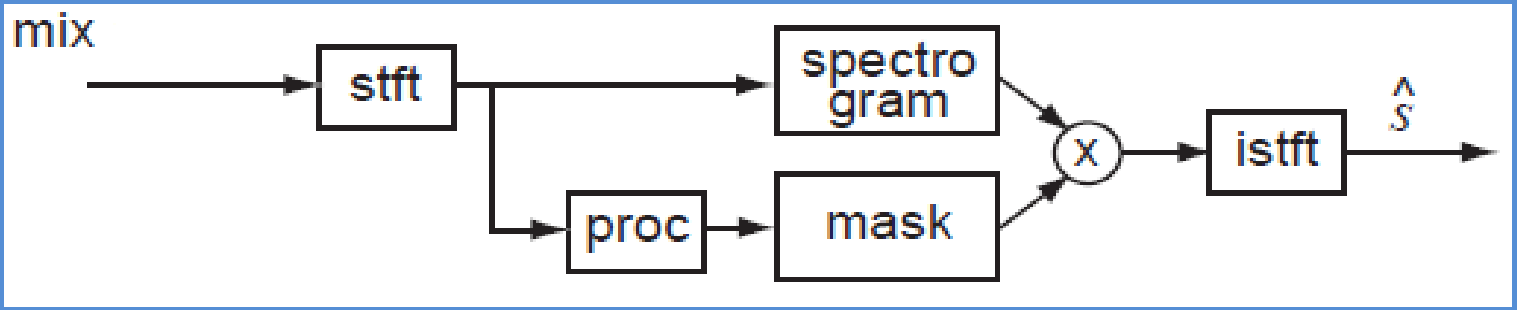


Linear to Nonlinear Separation

- Linear Separation: Multichannel ICA/IVA/Beamforming



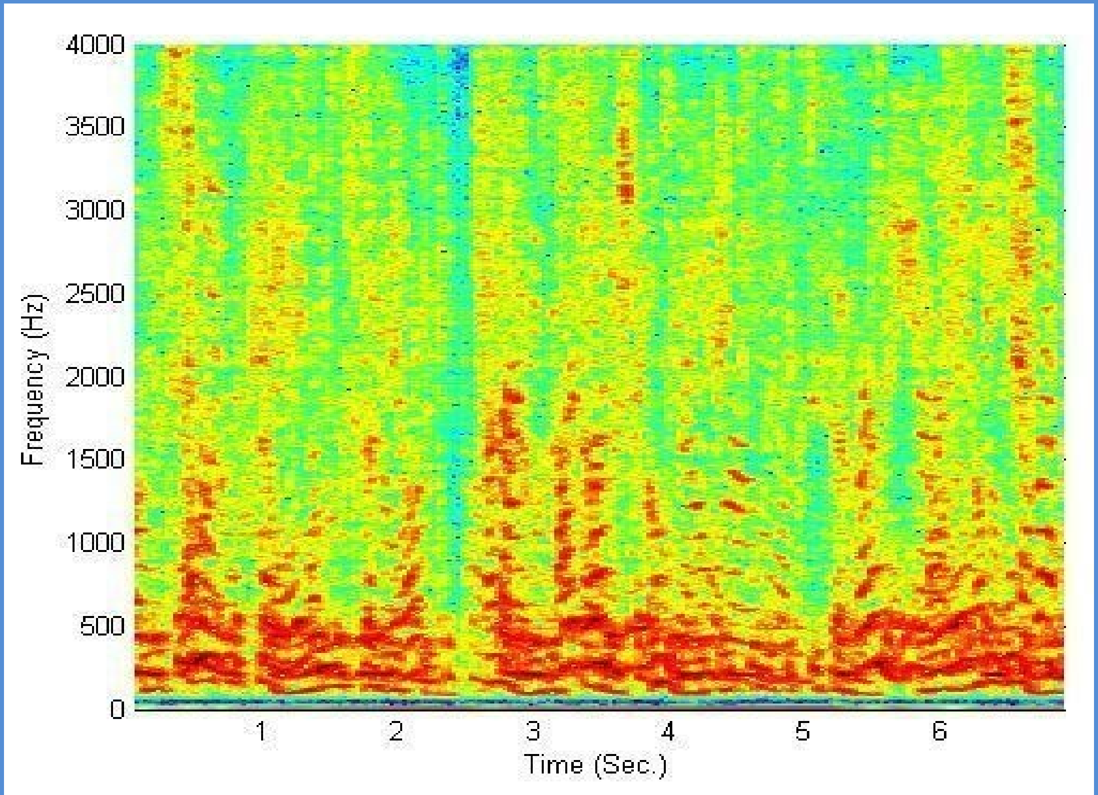
- Nonlinear Separation: Using a time frequency mask

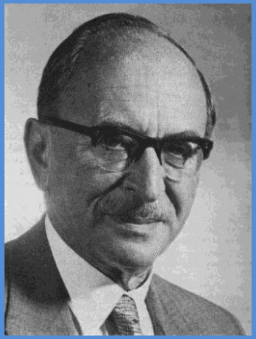


Time frequency masking ?

Time Frequency Signal Representation

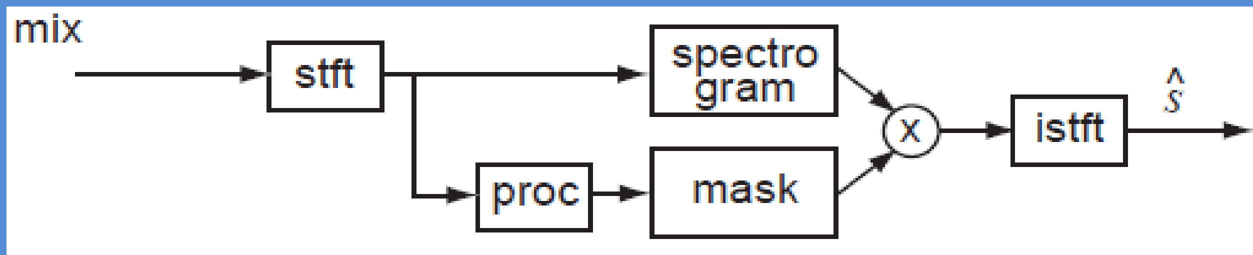
In 1946, Gabor proposed, “a new method of analysing signals is presented in which time and frequency play symmetrical parts”.

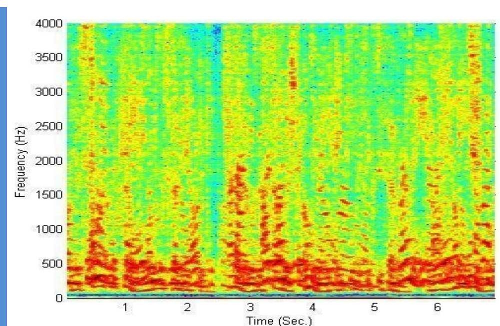




Time-Frequency Masking

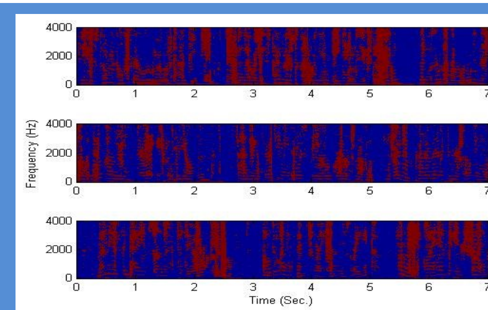
Audio signals are enhanced by simple nonlinear operations



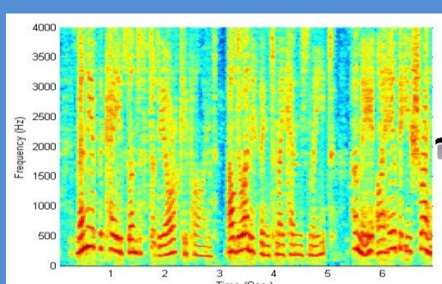


mixture

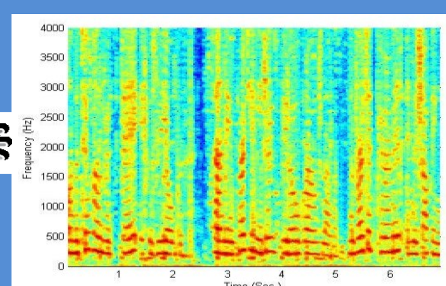
X



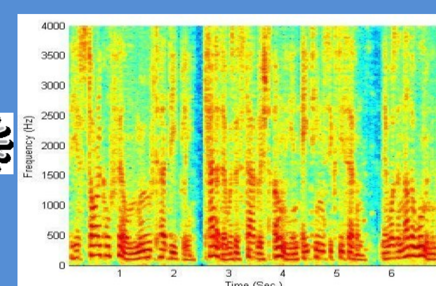
Masks



Source 1



Source 2



Source 3

Summary

In this talk, we have reviewed:

- Mathematical preliminaries
- BSS applications and concepts
- Sources and mixtures in BSS
- Representative block and sequential algos

You should be all set for the ensuing talks!

Acknowledgements

Jonathon Chambers wishes to express his 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.

His co-researchers: Dr Mohsen Naqvi and Mr Waqas Rafique.

Key Books and Reviews

- 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. Tickard, "Survey of Sparse and Non-Sparse Methods in Source Separation", *Int. Journal of Imaging Systems and Technology*, Vol.15, pp. 20-33, 2005.
- M. I. Mandel, R. J. Weiss, and D. P. W. Ellis, "Model-based expectation maximization source separation and localization," *IEEE Transactions on audio, speech, and language processing*, vol. 18, pp. 382–394, Feb. 2010.
- 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

References

- J.-F. Cardoso and Antoine Souloumiac. “Blind beamforming for non Gaussian signals”, In *IEE Proceedings-F*, 140(6):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 S.P.,* 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, March 2003, pp. 620-633.

- Aapo Hyvärinen, *“*Survey on Independent Component Analysis*”*, *Neural Computing Surveys*, Vol. 2, pp. 94-128, 1999.

References

- A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):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*, pages 757-763. MIT Press, Cambridge, MA, 1996.
- Parra, L. and Spence, C. (2000). Convolutive blind separation of non-stationary sources. *IEEE Trans. Speech Audio Processing*, 8(3):320–327.
- Te-Won Lee, Independent component analysis: theory and applications, Kluwer, 1998

References

- Buchner, H. Aichner, R. and Kellermann, W. (2004). 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*, pages 255–293. Kluwer Academic Publishers.
- Araki, S. Makino, S. Blin, A. Mukai, and H. Sawada, (2004). Underdetermined blind separation for speech in real environments with sparseness and ICA. In Proc. ICASSP 2004, volume III, pages 881–884.
- 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.
- Yi Hu; Loizou, P.C., "Evaluation of Objective Quality Measures for Speech Enhancement," Audio, Speech, and Language Processing, IEEE Transactions on , vol.16, no.1, pp.229,238, Jan. 2008Hiroe, A. (2006).