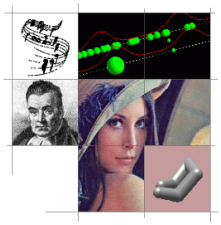
Blind Audio Source Separation

Vincent Y.F. Tan
yfvt2@cam.ac.uk

June 7, 2005



http://www2.eng.cam.ac.uk/~yfvt2/

Structure of Talk

- **Introduction and Preview.**
- **Blind Source Separation and Sparsity.**
- **Orthonormal Bases.**
- **Overcomplete Dictionaries.**
- **Results and Demonstration.**
- **Conclusions and Perspectives.**

Introduction

- **Consider the linear, instantaneous and noisy model**

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t + \mathbf{n}_t, \qquad 0 \leq t \leq N-1 \tag{1}$$

  - $\mathbf{x}_t = (x_{1,t}, \ldots, x_{m,t})^T$ **are the observations.**
  - $\mathbf{s}_t = (s_{1,t}, \ldots, s_{n,t})^T$ **are the sources**
  - $\mathbf{A} \in \mathbb{R}^{m \times n}$ **is a full rank mixing matrix and possibly** $m < n$**.**

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N}. \tag{2}$$

- **Task is to recover** $\mathbf{A}$**,** $\mathbf{S}$ **and** σ **given** $\mathbf{X}$**.**

Applications

- **Image Denoising and Compression.**
- **Telecommunications.**
- **Financial Time Series.**
- **Biomedicine - Foetus’ and Mother’s heartbeats.**

BSS: Method 1

- **Orthonormal Transform with analysis operator** $\boldsymbol{\Psi} \in \mathbb{R}^{N \times N}$

$$\mathbf{X}\boldsymbol{\Psi} = \mathbf{A}\mathbf{S}\boldsymbol{\Psi} + \mathbf{N}\boldsymbol{\Psi} \iff \tilde{\mathbf{X}} = \mathbf{A}\tilde{\mathbf{S}} + \tilde{\mathbf{N}}. \tag{3}$$

- **Assumptions**

1. **Sources follow Student** t **distribution in transformed domain.**
2. **Sources are independent** $p(\tilde{\mathbf{S}}) = \prod_{i=1}^{n} p(\tilde{\mathbf{s}}_i)$**.**
3. **Noise** $\mathbf{n}_t$ **is Gaussian i.i.d. with covariance** $\sigma^2 \mathbf{I}_m$

BSS: Method 2

The Gibbs Sampler estimates $\tilde{\mathbf{S}}$, $\mathbf{A}$ and σ and the hyperparameters.

Initialize $\boldsymbol{\theta}^{(0)} = \{\tilde{\mathbf{S}}^{(0)}, \mathbf{A}^{(0)}, \sigma^{(0)}, \mathbf{V}^{(0)}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\lambda}^{(0)}\}$.
for $k = 1 : K$ **do**

$$\tilde{\mathbf{S}}^{(k)} \sim p(\tilde{\mathbf{S}}|\mathbf{A}^{(k-1)}, \sigma^{(k-1)}, \mathbf{V}^{(k-1)}, \tilde{\mathbf{X}}) \quad (4)$$

$$\sigma^{(k)} \sim p(\sigma|\tilde{\mathbf{S}}^{(k)}, \tilde{\mathbf{X}}) \quad (5)$$

$$\mathbf{A}^{(k)} \sim p(\mathbf{A}|\tilde{\mathbf{S}}^{(k)}, \sigma^{(k)}, \tilde{\mathbf{X}}) \quad (6)$$

$$\mathbf{V}^{(k)} \sim p(\mathbf{V}|\tilde{\mathbf{S}}^{(k)}, \boldsymbol{\alpha}^{(k-1)}, \boldsymbol{\lambda}^{(k-1)}) \quad (7)$$

$$\boldsymbol{\alpha}^{(k)} \sim p(\boldsymbol{\alpha}|\mathbf{V}^{(k)}, \boldsymbol{\lambda}^{(k-1)}) \quad (8)$$

$$\boldsymbol{\lambda}^{(k)} \sim p(\boldsymbol{\lambda}|\mathbf{V}^{(k)}, \boldsymbol{\alpha}^{(k)}) \quad (9)$$

end for

BSS: Method 3

- **Reconstruct to obtain $\hat{\mathbf{S}}$ and $\hat{\mathbf{A}}$ and $\hat{\sigma}$.**

$$\tilde{\mathbf{X}} = \hat{\mathbf{A}}\hat{\tilde{\mathbf{S}}} + \hat{\tilde{\mathbf{N}}}, \tag{10}$$

$$\hat{\mathbf{S}} = \hat{\tilde{\mathbf{S}}}\boldsymbol{\Psi}^{-1} = \hat{\tilde{\mathbf{S}}}\boldsymbol{\Psi}^{T}. \tag{11}$$

- **Evaluate the performance using**

1. **Source to Distortion Ratio (SDR) - global criterion**
2. **Source to Interference Ratio (SIR)**
3. **Source to Artifacts Ratio (SAR)**
4. **Source to Noise Ratio (SNR)**

BSS: Method 4

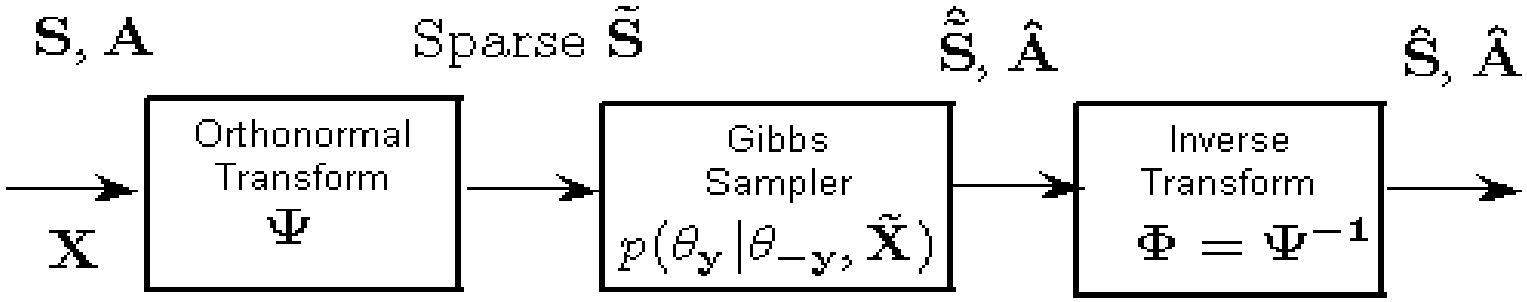


Figure 1: Block Diagram Summary of the BSS algorithm

Sparsity

- **Sources must have a sparse representation before sampling.**

  A signal is said to be sparse on a dictionary if only 'a few' coefficients of its decomposition are 'significantly' different from zero.

- **We use a sparsity index ξ to quantify sparsity.**

$$\xi \triangleq \frac{\|\tilde{\mathbf{s}}_i\|_1}{\|\tilde{\mathbf{s}}_i\|_2}. \tag{12}$$

- **The smaller ξ is, the sparser the signal $\tilde{\mathbf{s}}_i$.**

Orthonormal Transforms

1. **Discrete Cosine Transform (DCT)**
2. **Modified Discrete Cosine Transform (MDCT)**
3. **Discrete Wavelet Transform - Vaidyanathan (WT-Vai)**
4. **Discrete Wavelet Transform - Symmlet 8 (WT-Sym)**
5. **Wavelet Packet Best Basis (WPBB) on** $\mathbf{x}_1$
6. **No Transform or Standard Orthonormal Basis (NT)**

Overcomplete Dictionaries

1. **Short-Time Discrete Cosine Transform (STDCT)**
   - **Tapering overlapping windows.**
   - **25% overlap.**

2. **Hybrid Transform (HT)**
   - **Union of MDCT and wavelet basis.**

$$s(t) = \underbrace{s_{ton}(t)}_{\textbf{MDCT}} + \underbrace{s_{tr}(t)}_{\textbf{DWT}} + \underbrace{s_r(t)}_{\approx 0}, \tag{13}$$

   - **To model tonal and transient components of audio signals.**

‘Tonal vs Transient’ Balance

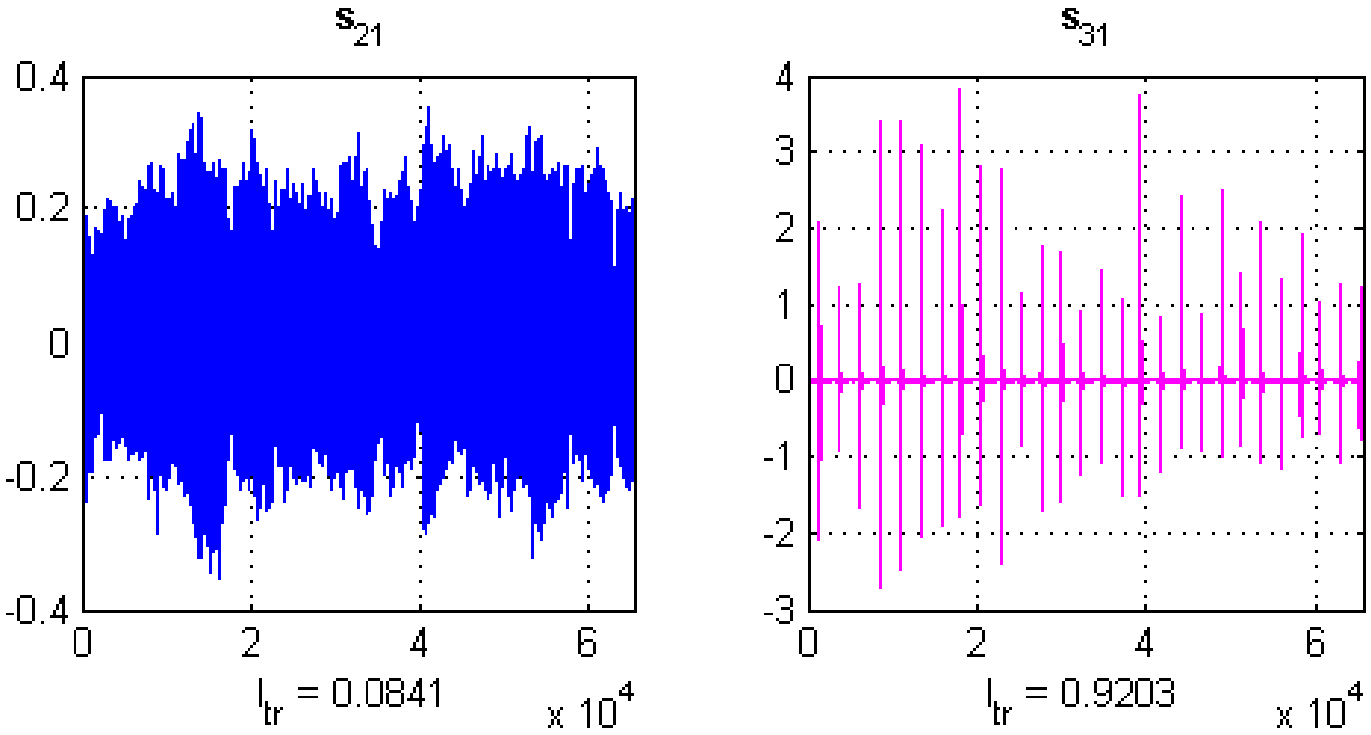


Figure 2: Tonal (Musical) and Transient (Percussion)

Results 1

- $n = 3$ **audio sources and** $m = 2$ **mixtures** $\Longrightarrow$ **underdetermined.**

- **Audio sources included:**

  1. $n = 3$ **speech signals**
  2. $n = 3$ **musical signals**
  3. $n = 3$ **percussion signals**
  4. **Combination of 1 speech, 1 musical and 1 percussion signal**

- **Calculated sparsity indices** ξ **and performance indices SDR, SIR, SAR and SNR for all transforms on all the sets of signals.**

Results 2

- The MDCT performed the best for the speech, musical and combination of signals.
  - Because these signals contain more tonals than transients.
- The DWT performed the best for the percussion signals.
  - Because percussion signals contain more transients.
- The WPBB finds a basis that performs well for all the signals.
  - Because it is adaptive.
- Performance depends highly on sparsity.

Results 3

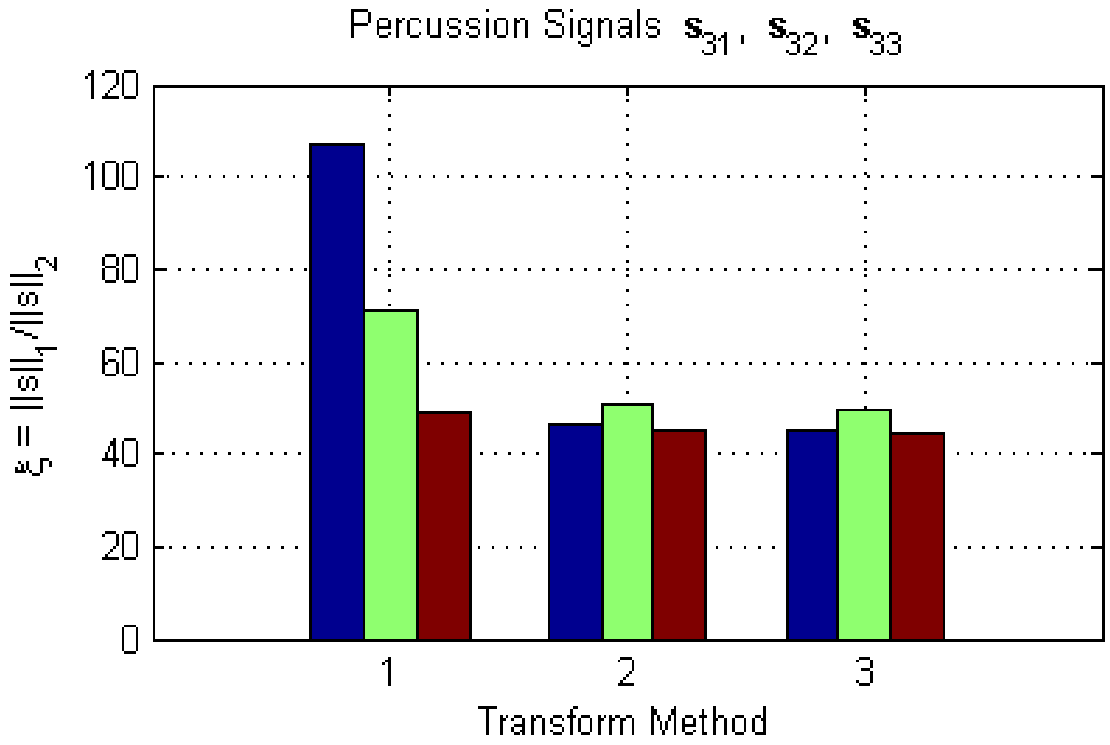


Figure 3: ξ for percussion signals; 1-MDCT, 2-WT-Vai, 3-WPBB

Results 4

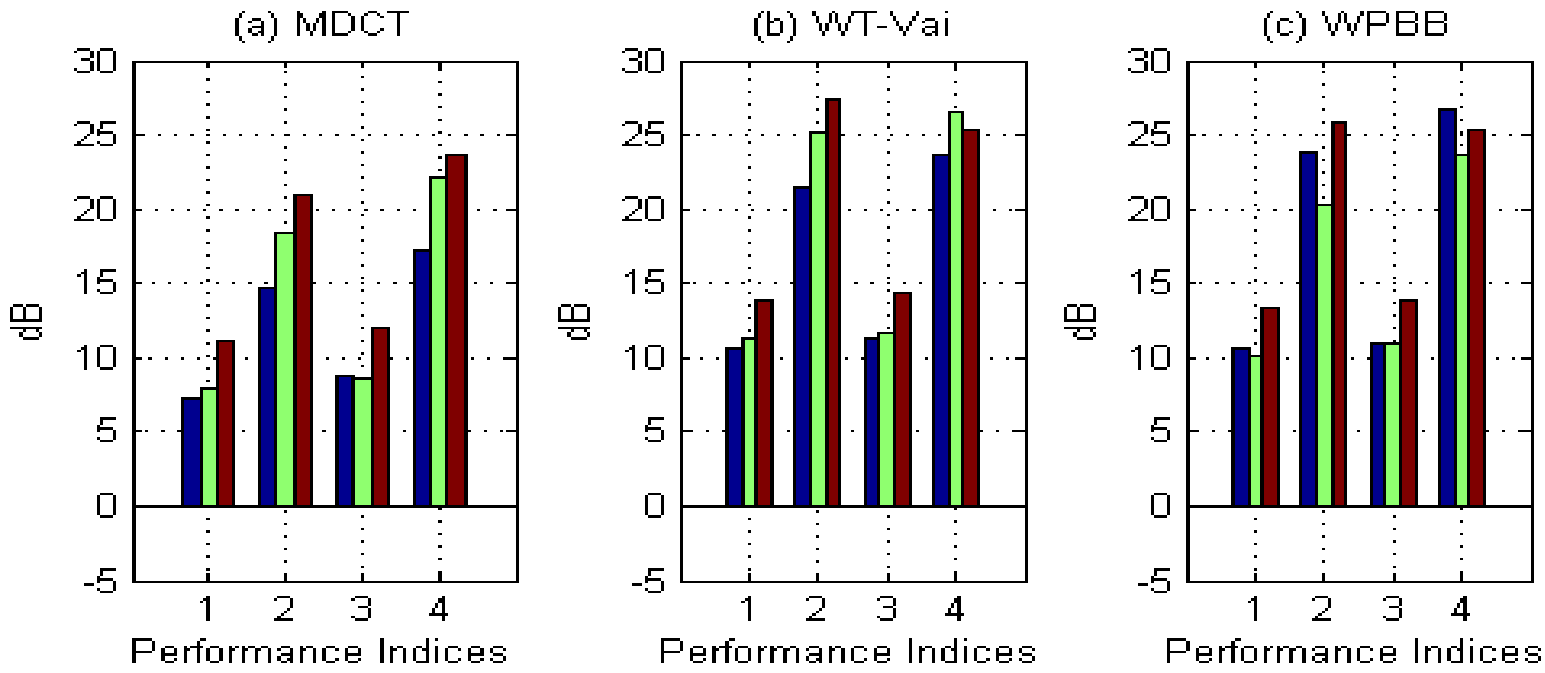


Figure 4: Performance Indices for percussion signals; 1-SDR, 2-SIR, 3-SAR, 4-SNR

Demonstration

Type	Musical			Percussion		
Sources	s_{21}	s_{22}	s_{23}	s_{31}	s_{32}	s_{33}

Type	Musical		Percussion	
Mixtures	x_1	x_2	x_1	x_2

Type	Musical			Percussion		
MDCT (SDR)	6.5	10.4	7.4	7.2	7.9	11.2
WT-Vai (SDR)	-0.3	4.6	1.0	10.7	11.3	13.8

Conclusions

- **'Tonal vs Transient' balance determines which transform performs the best.**
- **Best Basis Algorithm finds an excellent basis for all signals.**
- **Sparsity is very important.**
- **Overcomplete dictionaries provided marginal improvement.**
- **Heavy computational complexity.**

Conclusions

- **Thank you.**
- **All my demos can be found at my BSS homepage**
  http://www2.eng.cam.ac.uk/~yfvt2/bss_demo.html
- **Questions**