Monaural Source Separation

Y. Litvin

Department of Electrical Engineering

Supervised by Prof. Israel Cohen and Dr. Dan Chazan

Outline

1. Introduction
2. Subband Frequency Modulating Signal Modeling
3. Spectral Kurtosis
4. Bark-Scaled WPD
5. Conclusion

Outline

1. Introduction
   - Blind Source Separation
   - GMM Based Source Separation Algorithm
   - Distortion Measures
2. Subband Frequency Modulating Signal Modeling
3. Spectral Kurtosis
4. Bark-Scaled WPD
5. Conclusion

Blind Source Separation

Definition

Task of recovering a set of signals from a set of observed signal mixtures

- Number of sources
- Mixing model (instantaneous, echoic, convolutive, linear, non-linear)
- Number of observed mixtures
- Noise presence
- Training database

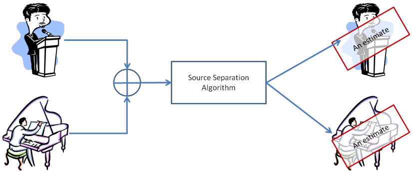
Problem Formulation

- Problem setup: single observation, two audio sources (speech and background music), no noise

$$x(n) = s_1(n) + s_2(n)$$

- In the STFT domain (benefits: low inter-band correlation, sparse representation, binary masks)

$$X_k(m) = S_{1,k}(m) + S_{2,k}(m)$$



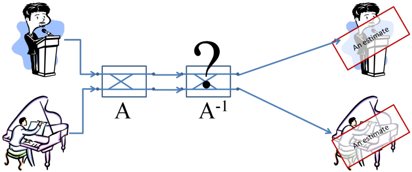
Previous Work
Multichannel

(Comon, 1994)

Find demixing matrix by minimizing some measure of statistical independence (ICA).

(Zibulevsky & Pearlmutter, 2001)

Find demixing matrix by minimizing some measure of sparsity.



Previous Work

Single channel

(Hanson & Wong, 1984)

Estimate pitch of one of the talkers. Used harmonic information and spectral subtraction to suppress the other.

(Bach & Jordan, 2005)

Define distances between each T-F bins using CASA principles. Use clustering to group similar T-F bins together. Apply binary masking in the T-F domain.

Previous Work

Single channel

(Virtanen, 2003)

Use Non-negative Matrix Factorization to factor spectral magnitude matrix into frequency basis vectors and amplitudes: $\mathbf{X} \approx \mathbf{AS}$. Cluster frequency basis vectors (columns of $\mathbf{A}$) and recreate mixture components using its frequency basis.

Wiener Based BSS Using GMM

Signal Model

- Introduced in (Benaroya & Bimbot, 2003)
- Mixture components are

$$\mathbf{s}_1 \sim N(0, \Sigma_1)\,; \mathbf{s}_2 \sim N(0, \Sigma_2)$$

- Observed signal is

$$\mathbf{x} = \mathbf{s}_1 + \mathbf{s}_2$$

- Posterior Mean (PM) estimator for $s_1(n)$ is

$$\hat{\mathbf{s}}_1 = \Sigma_1 (\Sigma_1 + \Sigma_2)^{-1} \mathbf{x}$$

Wiener Based BSS Using GMM

Signal Model in Fourier Domain

- Assume $\mathbf{s}_1, \mathbf{s}_2$ are stationary and approximately circular then Fourier transform $\mathscr{F}$ diagonalizes covariance matrix
- Denote $\mathbf{X} \triangleq \mathscr{F}\mathbf{x}, \mathbf{S}_1 \triangleq \mathscr{F}\mathbf{s}_1, \mathbf{S}_2 \triangleq \mathscr{F}\mathbf{s}_2$

$$\mathbf{S}_1 \sim N\left(0, \operatorname{diag}\left(\sigma_1^2\right)\right); \mathbf{S}_2 \sim N\left(0, \operatorname{diag}\left(\sigma_2^2\right)\right)$$
$$\mathbf{X} = \mathbf{S}_1 + \mathbf{S}_2$$

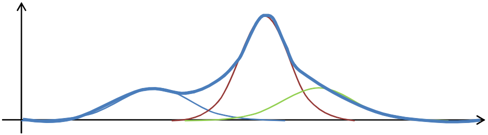
- PM estimator for the case of vectors with diagonal covariance matrix

$$\hat{\mathbf{S}}_1(i) = \frac{\sigma_1^2(i)}{\sigma_1^2(i) + \sigma_2^2(i)} \mathbf{X}(i)$$

Wiener Based BSS Using GMM

Gaussian Mixture Model

- Assume K Gaussian distributions $\left\{\mu^{(k)}, \Sigma^{(k)}\right\}_{k=1}^{K}$
- Probability of selecting k-th distribution is ω_k $(\sum_{k=1}^{k} \omega_k = 1)$
- GMM model defined by $\Lambda = \left\{\omega_k, \mu^{(k)}, \Sigma^{(k)}\right\}_{k=1}^{K}$



Wiener Based BSS Using GMM

Separation

- Assume $\mathbf{S}_c(m)$ generated by Λ_c ($c \in \{1,2\}$ class index)
- Introduce hidden variables $q_c \in \{1,\ldots,K\}$
- Define posterior probability $\gamma_{j,k} = p(q_1 = j, q_2 = k|\mathbf{X})$
- When conditioned on q_1, q_2, mixture components $\mathbf{S}_c \sim N\left(\mu^{(q_c)}, \Sigma^{(q_c)}\right)$ and we may use PM

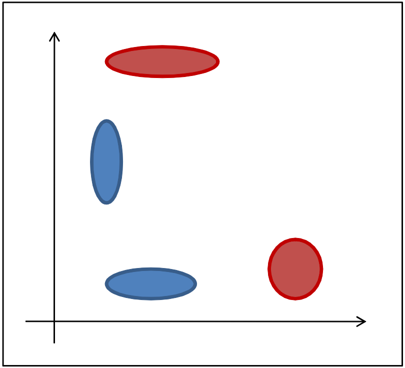
$$\hat{\mathbf{S}}_1(i) = \sum_{i,j} \gamma_{i,j} \frac{\sigma_1^{(i)2}(i)}{\sigma_1^{(i)2}(i) + \sigma_2^{(j)2}(i)} \mathbf{X}(i)$$

- $\gamma_{j,k}$ estimated from mixture observation by exhaustive enumeration of $j,k \in \{1,\ldots,K\}$

$$\begin{aligned} \gamma_{i,j} &\propto p(\mathbf{X}|q_1 = j, q_2 = k)\, p(q_1 = j)\, p(q_2 = k) \\ &= g\left(\mathbf{X}; \Sigma_1^{(j)} + \Sigma_2^{(k)}\right) w_1^{(j)} w_2^{(k)} \end{aligned}$$

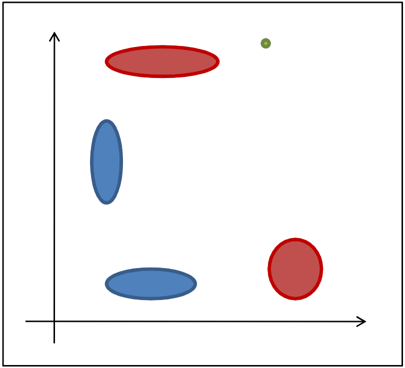
Wiener Based BSS Using GMM

Separation



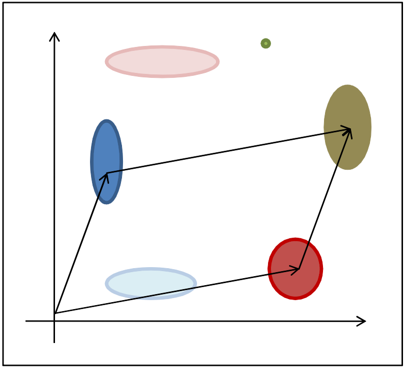
Wiener Based BSS Using GMM

Separation



Wiener Based BSS Using GMM

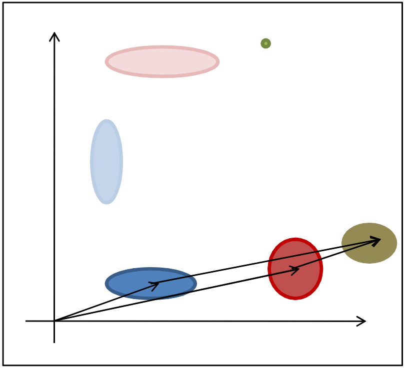
Separation



$\gamma_{i,j}$ is small

Wiener Based BSS Using GMM

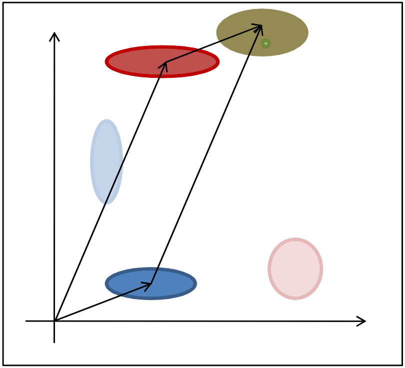
Separation



$\gamma_{i,j}$ is small

Wiener Based BSS Using GMM

Separation



$\gamma_{i,j}$ is large

Distortion Measures

- s_c the desired source, $s_{c'}$ the interfering source (Gribonval et al., 2003)

$$\hat{s}_c = y_c + e_{c,\text{interf}} + e_{c,\text{artif}} \qquad y_c := \langle \hat{s}_c, s_c \rangle \, s_c$$

$$e_{c,\text{artif}} := \hat{s}_c - (y_c + \langle \hat{s}_c, s_{c'} \rangle \, s_{c'}) \qquad e_{c,\text{interf}} := \langle \hat{s}_c, s_{c'} \rangle \, s_{c'}$$

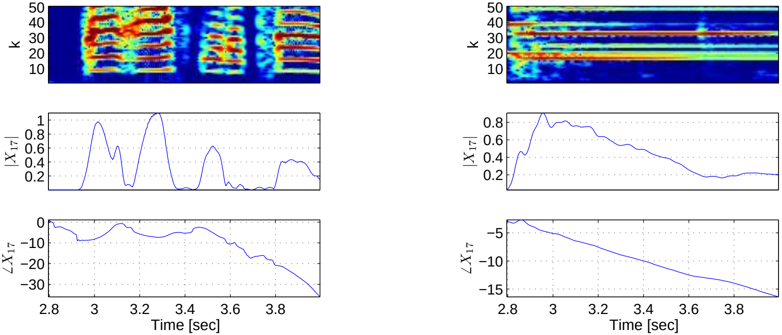
Distortion Measure	Definition
Signal to Distortion Ratio (SDR)	$10 \log_{10} \frac{\|y_c\|^2}{\|e_{c,\text{interf}} + e_{c,\text{artif}}\|^2}$
Signal to Interference Ratio (SIR)	$10 \log_{10} \frac{\|y_c\|^2}{\|e_{c,\text{interf}}\|^2}$
Signal to Artifact Ratio (SAR)	$10 \log_{10} \frac{\|y_c + e_{c,\text{interf}}\|^2}{\|e_{c,\text{artif}}\|^2}$

Outline

1. Introduction
2. Subband Frequency Modulating Signal Modeling
   - Motivation
   - AM-FM Demodulation using DESA
   - Energy of Frequency Modulating Signal
   - Source Separation Algorithm
   - Experimental Results
3. Spectral Kurtosis
4. Bark-Scaled WPD
5. Conclusion

Motivation

- Pitch track behavior is very different for speech and some "mechanically" generated sounds (e.g. music).
- Easily detected by examining the unwrapped phase of the subband signal



Teager's Energy Tracking Operator

- Undriven linear undamped oscillator with an amplitude A

$$\begin{aligned} E_{\text{osc}} &= \frac{1}{2} m \dot{x}_c^2 + \frac{1}{2} k x_c^2 = \frac{1}{2} m (A\omega_0)^2 \\ \omega_o &= \sqrt{k/m} \end{aligned}$$

- Teager Energy Operator (Teager & Teager, 1985)

$$\Psi_c[x(t)] = (\dot{x}(t))^2 - x(t)\ddot{x}(t)$$

Body position

$$\begin{aligned} x(t) &= A\cos(\omega_0 t + \theta) \\ \Psi_c[x(t)] &= 2A^2\omega_0^2 \end{aligned}$$

Approximately holds also for $A(t)$ and $\omega_0(t)$ (Maragos et al., 1993)

$$\begin{aligned} x(t) &\approx A(t)\cos(\omega_0(t)t + \theta) \\ \Psi_c[x(t)] &\approx 2A(t)^2\omega_0(t)^2 \end{aligned}$$

Energy Separation Algorithm (ESA)

- Continuous Energy Separation Algorithm (ESA) (Maragos et al., 1993)

$$\omega_0(t) \approx \sqrt{\frac{\Psi_c[\dot{x}(t)]}{\Psi_c[x(t)]}}$$

$$|A(t)| \approx \frac{\Psi_c[x(t)]}{\sqrt{\Psi_c[\dot{x}(t)]}}$$

Discrete ESA (DESA)

- Discrete TEO

$$\Psi\left[x(n)\right] = x^2(n) - x(n-1)\,x(n+1)$$

- Discrete ESA (DESA)

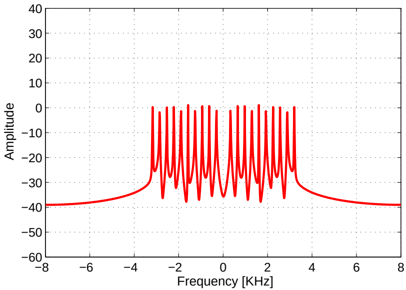
$$\hat{\Omega}_i(n) = \frac{1}{2}\arccos\left(1 - \frac{\Psi\left[x(n+1) - x(n-1)\right]}{2\Psi\left[x(n)\right]}\right)$$

$$|\hat{a}(n)| \approx \frac{2\Psi\left[x(n)\right]}{\sqrt{\Psi\left[x(n+1) - x(n-1)\right]}}$$

Energy of Frequency Modulating Signal
Harmonic signal

- Let x_ℓ be ℓ-th harmonic partial. Assume AM-FM model.

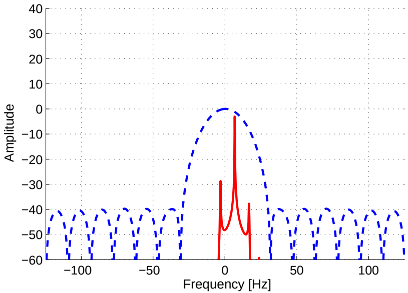
$$x(n) = \sum_\ell a_\ell(n)\cos\left(\Omega_0 \ell n + \sum_{i=0}^{n} r(i)\,\ell\frac{1}{T} + \theta_\ell\right)$$



Energy of Frequency Modulating Signal
STFT subband

- At the output of the STFT filterbank

$$\begin{aligned} X_k(m) &\approx a(mM)\, e^{j\left(\tilde{\Omega}_c mM + \sum_{i=0}^{mM} r(i)\frac{1}{T}\right)} \\ \tilde{\Omega}_c &= \Omega_c - \frac{2\pi}{N}k \end{aligned}$$

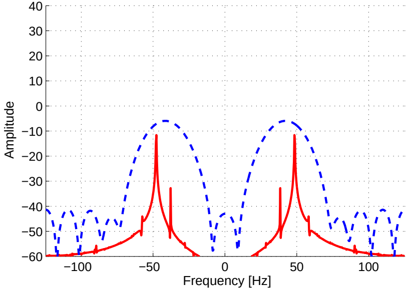


Energy of Frequency Modulating Signal

Intermediate Frequency

- Modulate subband to some intermediate frequency $\Omega_{\text{if}} = \alpha\pi$, $0 < \alpha < 1$

$$\tilde{X}_k(m) = \Re\left(X_k(m)\, e^{j\Omega_{\text{if}} m}\right) = a(mM)\cos\left(\left(\tilde{\Omega}_c + \Omega_{\text{if}}\right) mM + \sum_{i=0}^{mM} r(i)\,\frac{1}{T}\right)$$



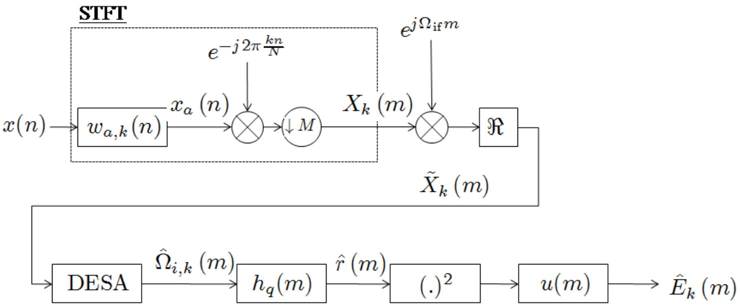
Energy of Frequency Modulating Signal
DESA

- Estimate instantaneous frequency using DESA

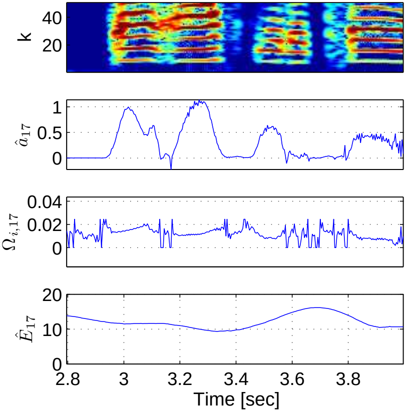
$$\begin{aligned}\hat{\Omega}_{i,k}(m) &\approx \frac{1}{2}\arccos\left(1-\frac{\Psi\left[\tilde{X}_k(m+1)-\tilde{X}_k(m-1)\right]}{2\Psi\left[\tilde{X}_k(m)\right]}\right)\\ &= \left(\tilde{\Omega}_c+\Omega_{\mathrm{if}}\right)M+r(mM)\frac{1}{T}\end{aligned}$$

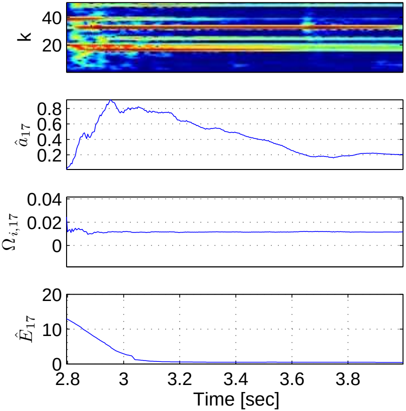
- Constant term is removed using high-pass filter h_r and Energy of Frequency Modulating Signal is obtained by smoothing $r^2(mM)$ using a Hamming window $u(m)$ of length N_u
- Upper bound on $M \leq \min\{\alpha N, (1-\alpha)N\}$ (due to DESA assumption on signal bandwidth)

Energy of Frequency Modulating Signal
Block Diagram



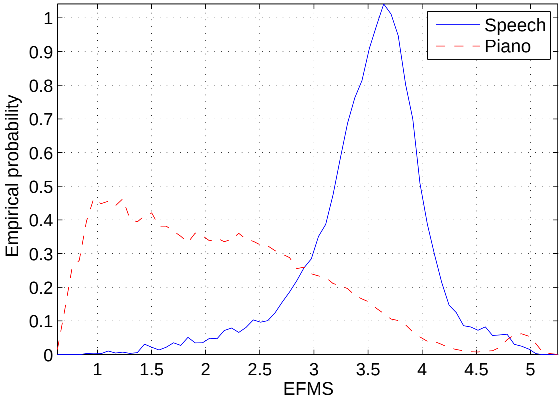
EFMS of Real Audio Signals





EFMS of Real Audio Signals

Probability distribution of EFMS



Source Separation Procedure
Classification

- ξ - EFMS value
- λ_{ij} - penalty for assigning a sample ξ to class i when in fact the sample belongs to class j
- λ_r - penalty for rejecting a sample
- We look for regions $\mathscr{R}_1, \mathscr{R}_2, \mathscr{R}_r$ that minimize loss function

$$L = \int_{\mathscr{R}_1} \lambda_{12} p\left(H^{(2)}|\xi'\right) p\left(\xi'\right) d\xi' +$$
$$+ \int_{\mathscr{R}_2} \lambda_{21} p\left(H^{(1)}|\xi'\right) p\left(\xi'\right) d\xi + \int_{\mathscr{R}_r} \lambda_r p\left(\xi'\right) d\xi'$$

Source Separation Procedure
Classification

- Let $\eta \triangleq \frac{p\left(\xi|H^{(1)}\right)p\left(H^{(1)}\right)}{p\left(\xi|H^{(2)}\right)p\left(H^{(2)}\right)}$
- Decision rules are

$$\xi \in \mathscr{R}_1 \iff \begin{cases} \frac{\lambda_{12}}{\lambda_{21}} < \eta \\ \frac{\lambda_r}{\lambda_{12}} > \frac{1}{1+\eta} \end{cases} \qquad \xi \in \mathscr{R}_2 \iff \begin{cases} \frac{\lambda_{12}}{\lambda_{21}} > \eta \\ \frac{\lambda_r}{\lambda_{21}} > \frac{1}{1+1/\eta} \end{cases}$$

$$\xi \in \mathscr{R}_r \iff \begin{cases} \frac{\lambda_r}{\lambda_{12}} \leq \frac{1}{1+\eta} \\ \frac{\lambda_r}{\lambda_{21}} \leq \frac{1}{1+1/\eta} \end{cases}$$

Source Separation Procedure

Masking

- Define binary mask for class $c \in \{1,2\}$

$$M_k^{(c)}(m) = \begin{cases} 1 & \xi_k(m) \in \mathscr{R}_c \\ 0 & \text{otherwise} \end{cases}$$

- Obtain masked STFT domain signal

$$\hat{X}_k^{(c)}(m) = M_k^{(c)}(m) X_k(m)$$

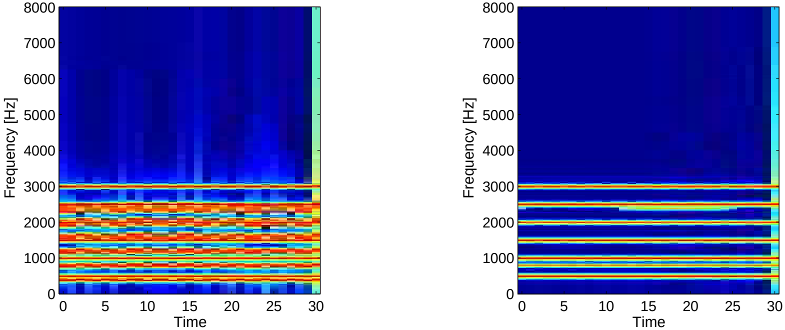
- Back to the time domain

$$\hat{x}^{(c)}(n) = \text{ISTFT}\left\{\hat{X}_k^{(c)}(m)\right\}$$

Experimental Results

Synthetic signals

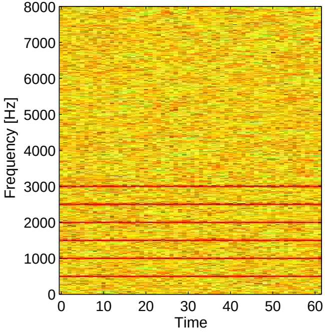
$$
\begin{aligned}
s_c(n) &= \sum_{\ell=0}^{N_h} \cos\left(\ell \cdot 2\pi f_c^{(c)} n/f_s + \sum_{m=0}^{n} q_\ell^{(c)}(m)\,\frac{1}{T}\right) \\
q_\ell^{(c)}(n) &= \ell \cdot d^{(c)} \cos\left(2\pi f_m^{(c)} n/f_s\right) \quad d^{(1)} = 20, d^{(2)} = 1
\end{aligned}
$$

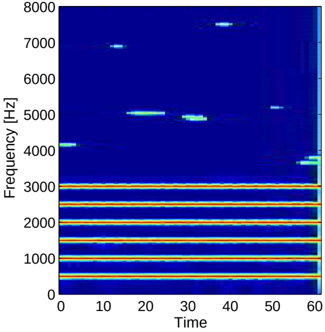


Experimental Results

Synthetic signals

- s_1 white noise, s_2 as before





Experimental Results

Real signals

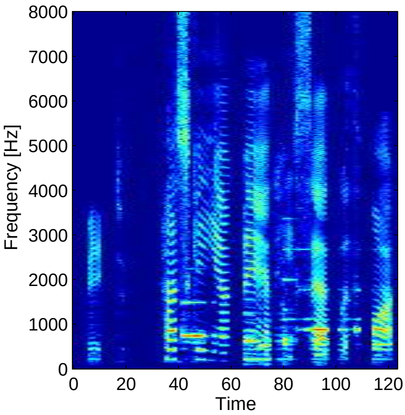
- Demo...
- $N = 1024,\ M = 64,\ N_u = 121,\ \delta_E = 15\text{dB},\ \lambda_{12} = \lambda_{21} = 1,\ \lambda_r = \infty,\ \alpha = 1/3$
- Oracle masks

$$\tilde{M}_k^{(c)}(m) = \begin{cases} 1 & |S_{1,k}(m)| \lessgtr |S_{2,k}(m)| \\ 0 & \text{otherwise} \end{cases}$$

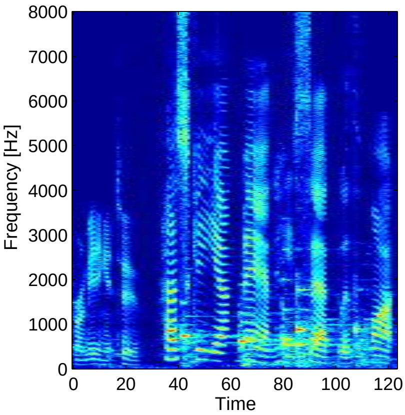
	SDR_1	SIR_1	SAR_1	LSD_1	SDR_2	SIR_2	SAR_2	LSD_2
Oracle mask	18.9	42.6	18.9	0.73	17.9	47.2	18.0	0.8
EFMS	**6.1**	**11.8**	**7.8**	**2.4**	**6.4**	**19.7**	**6.6**	**2.0**
GMM	2.4	9.3	3.8	2.9	2.6	7.9	4.8	2.5

Experimental Results

Real signals - Speech



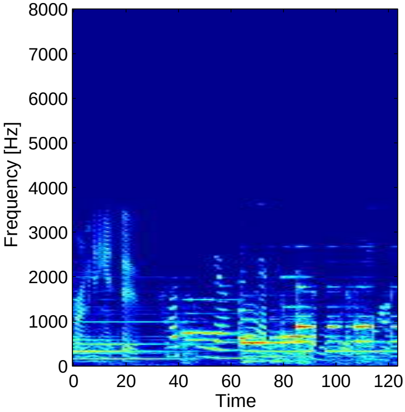
GMM



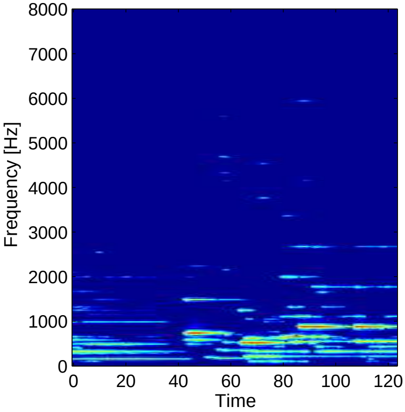
EFMS

Experimental Results

Real signals - Piano



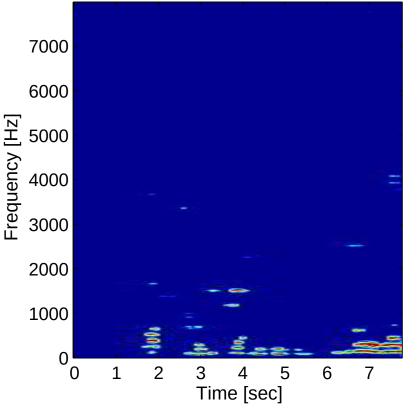
GMM



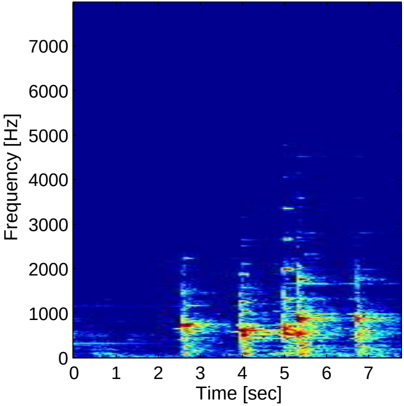
EFMS

Experimental Results

Real signals - Residual signal



Speech



Piano

Outline

1. Introduction
2. Subband Frequency Modulating Signal Modeling
3. Spectral Kurtosis
    - Kurtosis
    - SK of Audio Signals
    - Separation Algorithm
    - Experimental Results
4. Bark-Scaled WPD
5. Conclusion

Real Kurtosis

- Measure of peakedness
- Kurtosis definition

$$\mathrm{Kurt}(X) = \frac{\kappa_4}{\kappa_2^2} = \frac{\mathbb{E}\left(X^4\right)}{\mathbb{E}\left(X^2\right)^2} - 3$$

  - $X \sim N\left(\mu, \sigma^2\right) \Rightarrow \mathrm{Kurt}(X) = 0$
  - $X \sim \mathrm{Laplace}(\mu, b) \Rightarrow \mathrm{Kurt}(X) = 3$

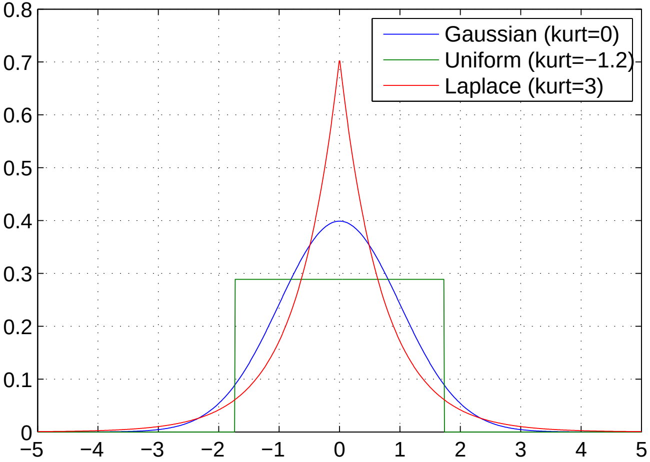
- Cumulant generating function of r.v. X

$$g(t) = \log \mathbb{E}\left(e^{tX}\right)$$

- k-th cumulant is given by

$$\kappa_k = \left. \frac{d^k}{dt^k} g \right|_{t=0}$$

Real Kurtosis



Peakiness of various distributions

Spectral Kurtosis Definition

- Let $x(n)$ be a time domain signal, X_k the k-th coefficient of DFT and X_k^* its complex conjugate. SK $\mathcal{K}_x$ is defined by (Vrabie et al., 2003)

$$\mathcal{K}_x(k) = \frac{\kappa\{X_k, X_k^*, X_k, X_k^*\}}{\left(\kappa\{X_k, X_k^*\}\right)^2}$$

- For circular processes

$$\mathcal{K}_x(k) = \frac{\mathbb{E}\left\{|X_k|^4\right\}}{\left(\mathbb{E}\left\{|X_k|^2\right\}\right)^2} - 2$$

  - $X \sim N(\mu, \sigma^2) \Rightarrow \mathrm{Kurt}(X) = 0$
  - $X = e^{j\Omega n + \theta}, \theta \sim U[0, 2\pi] \Rightarrow \mathrm{Kurt}(X) = -1$

Spectral Kurtosis of a Mixture

- Let $\phi_A(k) \triangleq \mathbb{E}\left(|A_k|^2\right)$
- Let $\gamma \triangleq \phi_{s_1}(k)/\phi_{s_2}(k)$

$$\begin{aligned}\mathcal{K}_x(k) &= \left|\frac{\phi_{s_1}(k)}{\phi_{s_1}(k)+\phi_{s_2}(k)}\right|^2 \mathcal{K}_{s_1}(k) + \left|\frac{\phi_{s_2}(k)}{\phi_{s_1}(k)+\phi_{s_2}(k)}\right|^2 \mathcal{K}_{s_2}(k) \\ &= \left|\frac{1}{1+1/\gamma}\right|^2 \mathcal{K}_{s_1}(k) + \left|\frac{1}{1+\gamma}\right|^2 \mathcal{K}_{s_2}(k)\end{aligned}$$

(Benesty, 2009)

- $\phi_{s_1}(k) \gg \phi_{s_2}(k) \Rightarrow \gamma \gg 1 \Rightarrow \mathcal{K}_x(k) \approx \mathcal{K}_{s_1}(k)$
- $\phi_{s_1}(k) \ll \phi_{s_2}(k) \Rightarrow \gamma \ll 1 \Rightarrow \mathcal{K}_x(k) \approx \mathcal{K}_{s_2}(k)$
- From W-DO, for each TF bin $\gamma \ll 1$ or $\gamma \gg 1$

SK Estimation

- Let $X_k(m)$ be k-th frequency band of the STFT filterbank
- Assume $X_k(m)$ quasi-stationary i.i.d. process
- L number of samples
- Spectral Kurtosis unbiased estimator (Vrabie et al., 2003)

$$\hat{\mathscr{K}}_x(k) = \frac{L}{L-1}\left[\frac{(L+1)\sum_{i=1}^{L}|X_k(i)|^4}{\left(\sum_{i=1}^{M}|X_k(i)|^2\right)^2} - 2\right]$$

Physical Interpretation

- Let $Y_k(m) = |X_k(m)|^2$
- SK can be rewritten as (Antoni, 2006)

$$\hat{\mathcal{K}}_X(k) \triangleq \frac{\left[\left\langle Y^2 \right\rangle_m - \left\langle Y \right\rangle_m^2\right]}{\left\langle Y \right\rangle_m^2} - 1$$

- Can be seen as normalized empirical variance with respect to time
- Similar to SK up to a constant additive element

STSK Estimation

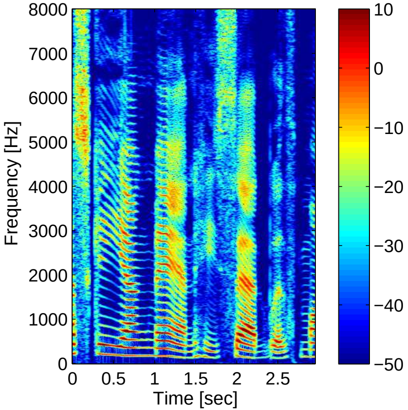
- Short Time Spectral Kurtosis (STSK) is a SK localized in time
- Let be $2n$-th moment estimator

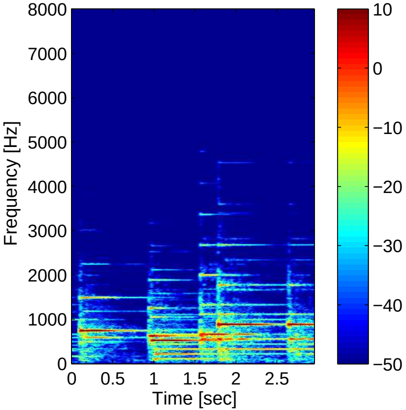
$$\hat{S}_{2nX,k}(m) \triangleq \sum_{i=-\lfloor L_K/2 \rfloor}^{\lfloor L_K/2 \rfloor} w_{\mathcal{K}}(m+i) \left| X_k(i) \right|^{2n}$$

- We define STSK

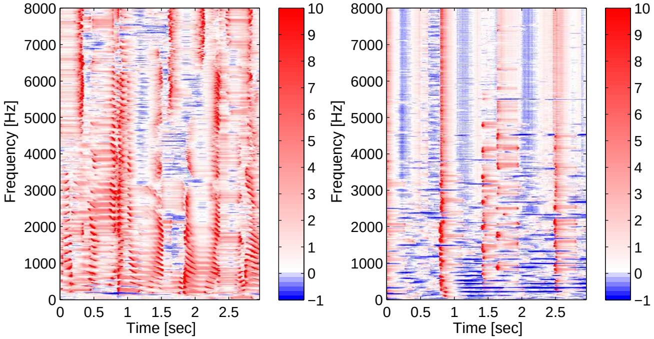
$$\hat{\mathcal{K}}_{X,k}(m) \triangleq \frac{\hat{S}_{4X,k}(m)}{\hat{S}^2_{2X,k}(m)} - 2$$

Speech and Piano Spectrograms

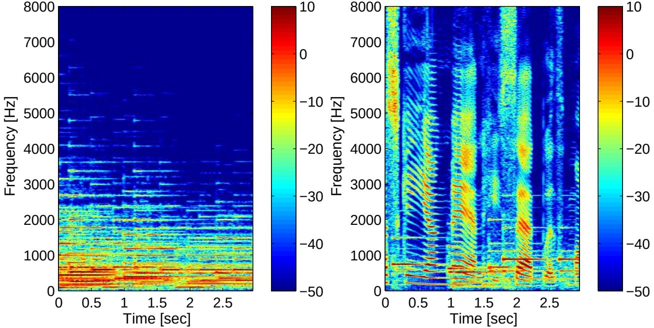




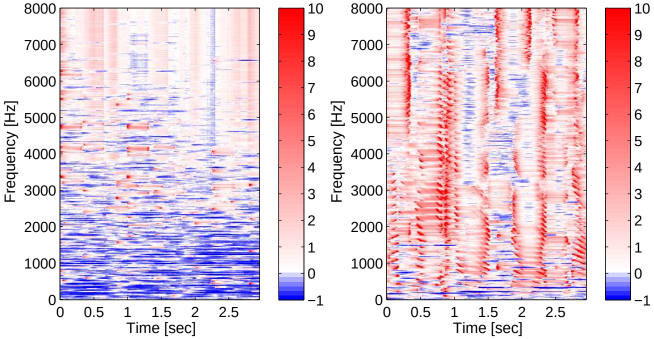
Speech and Piano STSK



Piano play (fast), Mix Spectrograms



Piano play (fast), Mix STSK



Separation

- Mask out time-frequency bins that belong to the interfering signal

$$M_{1,k}(m) = \begin{cases} 1 & \hat{\mathscr{K}}_x(m,k) > \delta_1 \\ 0 & \text{otherwise} \end{cases}$$

$$M_{2,k}(m) = \begin{cases} 1 & \hat{\mathscr{K}}_x(m,k) \leq \delta_2 \\ 0 & \text{otherwise} \end{cases}$$

- Recover desired signal ($\circ$ element-wise multiplication)

$$\hat{s}_c(n) = \mathrm{ISTFT}(M_c \circ X)$$

Experimental Results

- Demo ...

	SDR_1	SIR_1	SAR_1	$\mathbf{LSD_1}$	SDR_2	SIR_2	SAR_2	$\mathbf{LSD_2}$
Oracle mask	18.9	42.6	18.9	0.73	17.9	47.2	18.0	0.8
GMM	2.4	9.3	3.8	2.9	2.6	7.9	4.8	2.5
STSK	**7.7**	**19.5**	**8.0**	**2.8**	**8.2**	**21.3**	**8.4**	**2.3**
EFMS	6.1	11.8	7.8	2.4	6.4	19.7	6.6	2.0

Outline

1 Introduction

2 Subband Frequency Modulating Signal Modeling

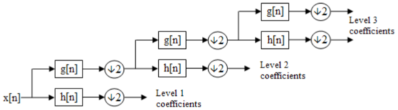
3 Spectral Kurtosis

4 Bark-Scaled WPD
- Algorithm
- Separation Algorithm
- Experimental Results

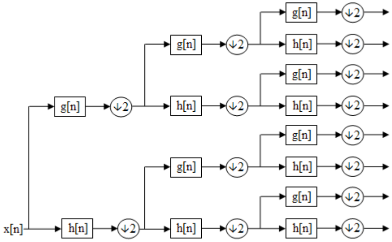
5 Conclusion

Wavelet Packet Decomposition

- Discrete Wavelet Transform



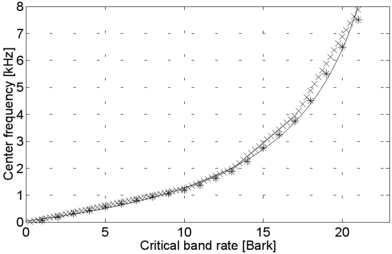
- Wavelet Packet Decomposition



Bark scale

- Basilar membrane acts as non uniform filterbank
- Accounts for non uniform frequency sensitivity of human year
- Bark scale follows center frequencies of critical bands (1 Bark apart)
- Frequency to Bark scale $z = 26.81/(1+1980/f) - 0.53$

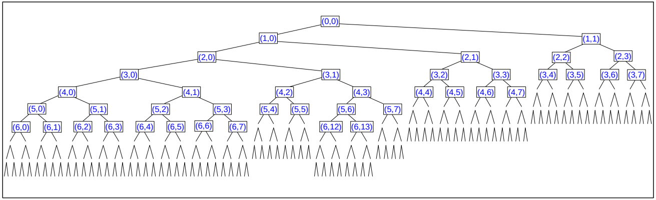
- Let L be depth of the WPD tree and $0 \leq l < L, 0 \leq n < 2^l$
- Center frequency of WPD node (l,n) is $f_{l,n} = 2^{-l}\left(\mathrm{GC}^{-1}(n) + 0.5\right)\frac{F_s}{2}$



from I. Cohen 2001

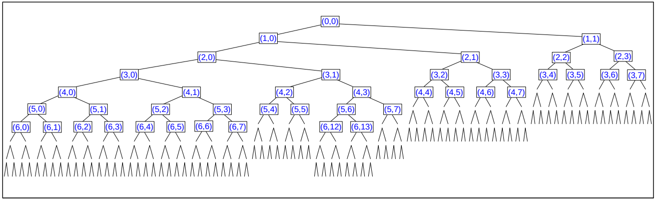
Bark-Scaled WPD

- Bark-Scaled WPD (BS-WPD) introduced in (Cohen, 2001)
- WPD with center frequencies located 1-Bark apart
- Critical band structured filterbank:
  - fine frequency resolution at low frequencies
  - coarse frequency resolution at high frequencies
- Various wavelet families may be used
- Improved frequency resolution by additional levels of decomposition



Constant Sampling Rate BS-WPD

- BS-WPD has different sampling rates at terminal nodes
- Stop decimating for nodes deeper than 6
- Total of 168 frequency bands comparing to 512 for STFT with similar bandwidth at low frequency bands



Mapping Based Complex Wavelet Transform

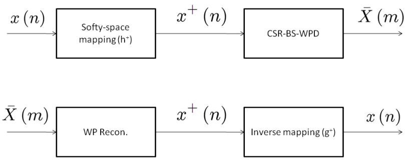
- DWT/WPD lack shift invariance
  - Two time domain signals $x(n), x_\Delta(n) = x(n-\Delta)$, small Δ
  - Let $X_{l,n}(m), X_{\Delta,l,n}(m)$ be (l,n) terminal node of DWT
  - $X_{l,n}(m)$ is significantly different from $X_{\Delta,l,n}(m)$
  - STFT transform: Δ mostly has influence on phase
- Reason: decimation in the decomposition tree

Mapping based Complex Wavelet Transform

- Introduced by (Fernandes et al., 2003)
- Achieves "approximate shiftability"
- Hardy-space $H^2(\mathbb{R} \to \mathbb{C})$ is defined by

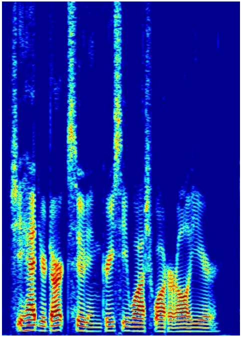
$$H^2(\mathbb{R} \to \mathbb{C}) \triangleq \{f \in L^2(\mathbb{R} \to \mathbb{C}) : \mathscr{F}f(\omega) = 0 \text{ for a.e. } \omega < 0\}$$

- $L^2(\mathbb{R} \to \mathbb{R})$ isomorphic to Hardy-space
- Softy-space is an approximation for a Hardy-space and can be mapped using digital filter h^+

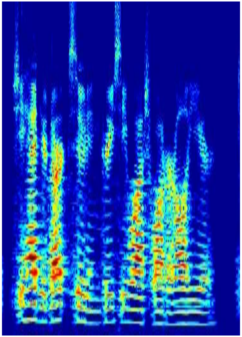


Time-Frequency Representation Comparison

STFT



Complex BS−WPD



Training

- $\mathbb{E}_m\left(\bar{S}_c\right) = 0 \Rightarrow \Lambda_c = \left\{\omega_k, 0, \Sigma^{(k)}\right\}_{k=1}^{K}$
- Data points $\left\{\bar{S}_1(m)\right\}_{m=1}^{L}, \left\{\bar{S}_2(m)\right\}_{m=1}^{L}$
- Using EM to train GMM models Λ_1, Λ_2

Separation

- Assume $\bar{S}_c(m)$ generated by Λ_c ($c \in \{1,2\}$ class index)
- Introduce variables $q_c \in \{1,\ldots,K\}$
- Define posterior probability $\gamma_{j,k} = p\left(q_1 = j, q_2 = k|\bar{X}\right)$
- When conditioned on q_1, q_2, mixture components $\bar{S}_c \sim N\left(\mu^{(q_c)}, \Sigma^{(q_c)}\right)$ and we may use PM

$$\hat{\bar{S}}_1(i) = \sum_{i,j} \gamma_{i,j} \frac{\sigma_1^{(i)2}(i)}{\sigma_1^{(i)2}(i) + \sigma_2^{(j)2}(i)} \bar{X}(i)$$

- $\gamma_{j,k}$ estimated from mixture observation by exhaustive enumeration of $j,k \in \{1,\ldots,K\}$

$$\begin{aligned}\gamma_{i,j} &\propto p\left(\bar{X}|q_1 = j, q_2 = k\right) p\left(q_1 = j\right) p\left(q_2 = k\right)\\ &= g\left(\bar{X}; \Sigma_1^{(j)} + \Sigma_2^{(k)}\right) w_1^{(j)} w_2^{(k)}\end{aligned}$$

Synthetic Signals

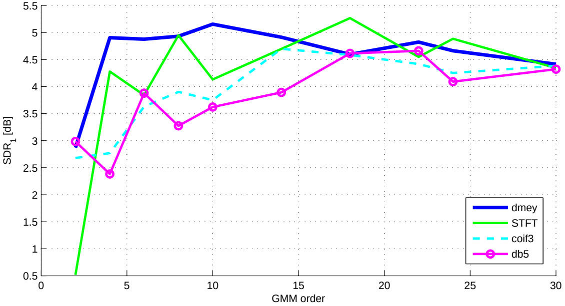
$$x_c(m) = \begin{cases} \sum_{i=1}^{2} \cos\left(\frac{2\pi}{f_s} f_{1,i}^{(c)} n\right) & w.p. \frac{1}{2} \\ \sum_{i=1}^{2} \cos\left(\frac{2\pi}{f_s} f_{2,i}^{(c)} n\right) & w.p. \frac{1}{2} \end{cases}$$

$$f_{1,1}^{(1)} = 220\text{Hz},\ f_{1,2}^{(1)} = 440\text{Hz},\ f_{1,1}^{(2)} = 300\text{Hz},\ f_{1,2}^{(2)} = 600\text{Hz}$$

	SDR_1	SIR_1	SAR_1	SDR_2	SIR_2	SAR_2
STFT	16	40	16	16	31	16
CSR-BS-WPD	20	35	20	22	40	22

Natural Signals

- Speech and piano play
- Compared to STFT based algorithm (Benaroya & Bimbot, 2003)
- Comparison parameters
  - GMM order
  - Wavelet family



Results Analysis

- Comparing to STFT
  - Low orders of GMM: better than STFT or comparable
  - High orders of GMM: comparable
- Different wavelet families: *dmey* superior to other wavelet families
- Approximate W-DO orthogonality (Yilmaz & Rickard, July 2004)
  - *dmey* CSR-BS-WPD transform has
    - the most sparse coefficients compared to other wavelet families
    - sparseness comparable to STFT
    - good frequency localization properties
    - successfully used for speech enhancement (Cohen, 2001)

Outline

1. Introduction
2. Subband Frequency Modulating Signal Modeling
3. Spectral Kurtosis
4. Bark-Scaled WPD
5. Conclusion

Summary
EFMS

- Definition of new signal analysis domain
  - Demonstration of usefulness in the task of source separation
- Novel monaural separation algorithm
- Based on subband phase signal properties (EFMS)
- Accounts for subband time dynamics and not spectral shape
- Good perceptual quality

Summary
STSK

- High order statistics (short time spectral kurtosis) for single channel source separation
  - Based on unpublished work of J. Benesty
  - Ad-hoc definition and estimator
  - Demonstration of usefulness in the task of source separation
- Defined STSK
- Like EFMS, provides good local TF signal characterization
- Study of STSK statistical properties is necessary
- Good experimental results

Summary
CSR-BS-WPD

- Extension of Bark-Scaled Wavelet Packet Decomposition (Cohen, 2001)
  - Approximate shiftability
  - Constant sampling rate
- Constant Sampling Rate Bark-Scaled signal analysis introduced
  - Critical band structure
  - Approximate shiftability
  - Easy access to spectral shape at given time index
- GMM based single channel source separation algorithm introduced
  - Reduced dimension of data points (compared to STFT)
  - Reduced computational complexity
  - Improved performance compared to STFT based algorithm

Future Research

- "Edge preserving" EFMS estimation (bilateral filtering?)
- More "sophisticated" FM analysis (e.g. spectral analysis of FM signal)
- Non-uniform filterbank
- Varying values EFMS for different frequencies
- Soft instead of binary masks
- Incorporate spectral information into classification
- Rigorous definition of STSK and its statistical properties
- Additional applications of EFMS, STSK and CSR-BS-WPD analysis (e.g. signal classification)

Thank you!

For Further Reading I

Antoni, Jèrôme. 2006.
The spectral kurtosis: a useful tool for characterising non-stationary signals.
Mechanical Systems and Signal Processing, **20**(2), 282 – 307.

Bach, F. R., & Jordan, M. I. 2005.
Blind One-microphone Speech Separation: A Spectral Learning Approach.
Pages 65–72 of: Saul, Lawrence K., Weiss, Yair, & Bottou, Léon (eds), *Advances in Neural Information Processing Systems 17*. Cambridge, MA: MIT Press.

For Further Reading II

Benaroya, L., & Bimbot, F. 2003 (Apr.).
Wiener Based Source Separation with HMM/GMM using a Single Sensor.
Pages 957–961 of: Proc. 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003).

Benesty, J. 2009 (Jul).
private communication.

Cohen, I. 2001.
Enhancement of speech using bark-scaled wavelet packet decomposition.
Pages 1933–1936 of: Eurospeech.

For Further Reading III

Comon, Pierre. 1994.
Independent component analysis, a new concept?
Signal Process., **36**(3), 287–314.

Fernandes, F.C.A., van Spaendonck, R.L.C., & Burrus, C.S. 2003.
A new framework for complex wavelet transforms.
IEEE Trans. Signal Process., **51**(7), 1825–1837.

Gribonval, R., Benaroya, L., Vincent, E., & Févotte, C. 2003 (Apr.).
Proposals for Performance Measurement in Source Separation.
Pages 763–768 of: Proc. 4th International Symposium on ICA and BSS (ICA2003).

For Further Reading IV

Hanson, B., & Wong, D. 1984 (Mar).
The harmonic magnitude suppression (HMS) technique for intelligibility enhancement in the presence of interfering speech.
Pages 65–68 of: Proc. Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-84, vol. 9.

Maragos, P., Kaiser, J.F., & Quatieri, T.F. 1993.
On amplitude and frequency demodulation using energy operators.
IEEE Trans. Signal Process., **41**(4), 1532–1550.

Teager, H. M., & Teager, S. M. 1985.
A phenomenological model for vowel production in the vocal tract.
Chap. 3, pages 73–109 of: Daniloff, R. G. (ed), *Speech Science: Recent Advances*.
San Diego, CA: College-Hill Press.

For Further Reading V

Vincent, E., Févotte, C., Benaroya, L., & Gribonval, R. 2003 (Apr.).
A Tentative Typology of Audio Source Separation Tasks.
Pages 715–720 of: Proc. 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003).

Virtanen, Tuomas. 2003.
Sound Source Separation Using Sparse Coding with Temporal Continuity Objective.
In: International Computer Music Conference, ICMC.

Vrabie, V., Granjon, P., & Servière, C.r. 2003.
Spectral Kurtosis: from Definition to Application.
In: IEEE-EURASIP International Workshop on Nonlinear Signal and Image Processing.

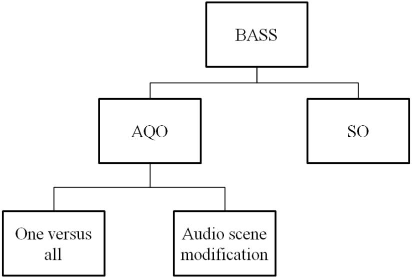
For Further Reading VI

Yilmaz, O., & Rickard, S. July 2004.
Blind separation of speech mixtures via time-frequency masking.
IEEE Trans. Signal Process., **52**(7), 1830–1847.

Zibulevsky, Michael, & Pearlmutter, Barak A. 2001.
Blind Source Separation by Sparse Decomposition in a Signal Dictionary.
Neural Comput., **13**(4), 863–882.

BASS Tasks Taxonomy

- Following taxonomy (Vincent et al., 2003)
- AQO - Audio quality oriented
  - One versus all
  - Audio scene modification
- SO - Significance oriented



Applications

- One versus all
  - track extraction from polyphonic music
  - speech enhancement
  - old recording restoration
  - karaoke
  - object-based audio coding
- Audio scene modification
  - remixing of existing recordings
  - signal enhancement in hearing aids
- Significance oriented
  - speaker identification
  - polyphonic music transcription
  - musical instrument identification in polyphonic music