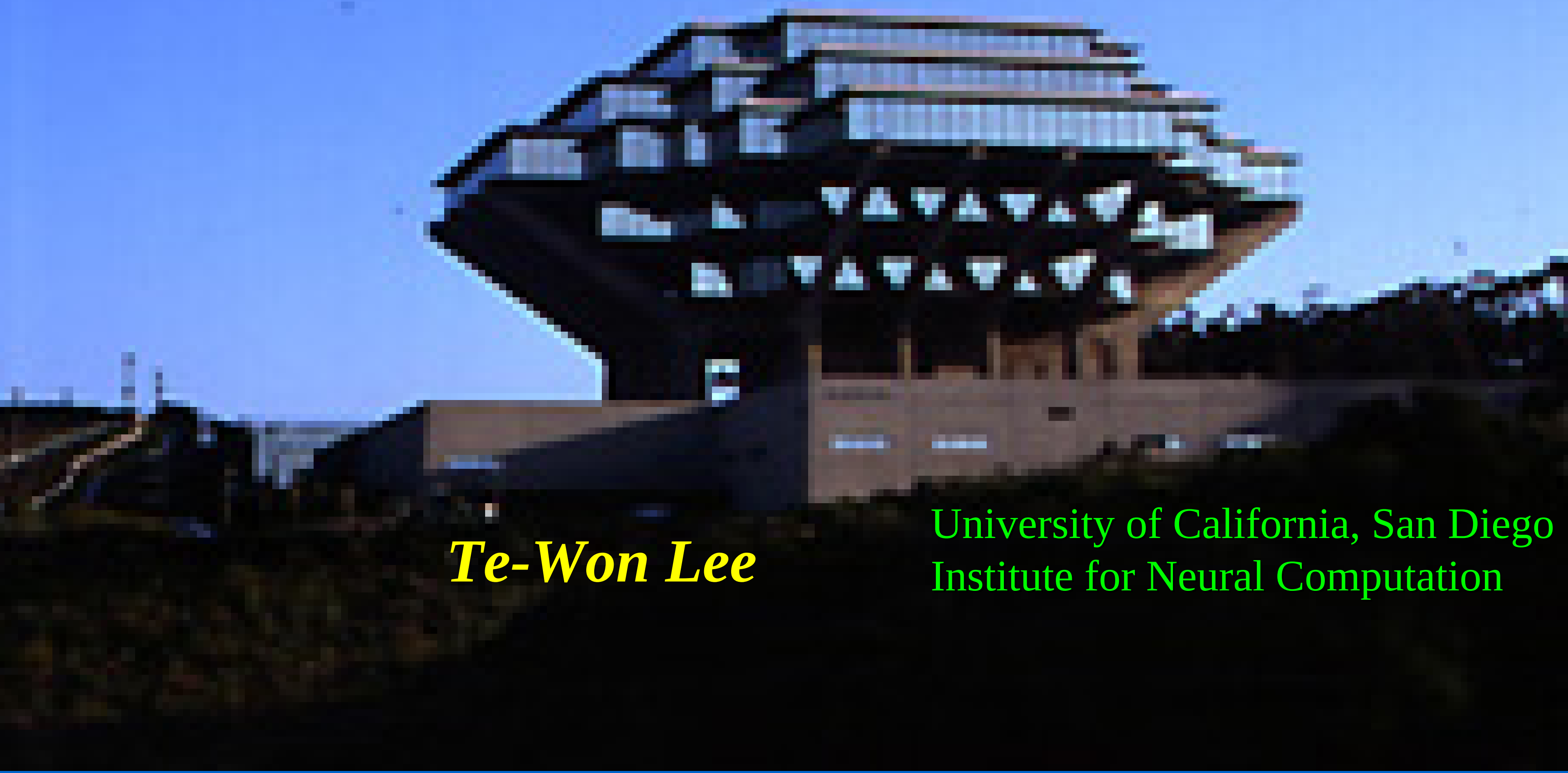
Blind Machine Separation



Te-Won Lee

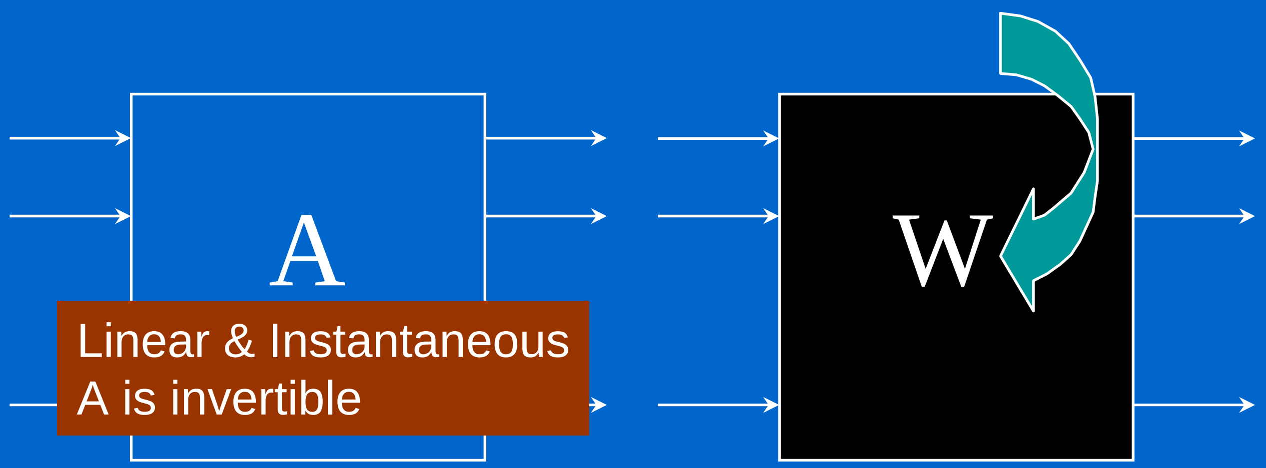
University of California, San Diego
Institute for Neural Computation

Blind Machine Separation

- Problem we want to solve:
  - Single microphone blind source separation & deconvolution
  - No prior on sources, mixing, dynamics
- Problem we can solve:
  - Blind source separation
    - Equal number of sources and sensors, no additive noise
    - Instantaneous and linear mixing, stationary and independent sources
- Problem we should solve:
  - Acoustic multichannel blind deconvolution:
    - Equal number of sources and sensors, no additive noise,
    - Convolved linear mixing, nonstationary and independent signals
- Problem we may solve:
  - Binaural Source Separation
    - Two sensors, multiple sources
    - Convolved linear mixing, strong source priors

Blind Source Separation

Evaluate cost function



s | $x=As$ | $u=Wx$

Sources | Observations | Recovered sources

Independent & stationary | No additive noise

- **Signal Processing**: Adaptive Filtering with nonlinear learning rule
- **Statistics**: Generative model with hidden variables: Synthesis & Analysis Model
- **Communications**: Encoding and decoding model for channel estimation

Independent Component Analysis

- **Essence:**
  - Contrast function that relates to mutual information & constraint optimization or gradient methods

- **Algorithms (classic) to standard ICA problem:**
  - Joint Approximate Diagonalization Equivariance (JADE) [Cardoso, 1993]
  - Cumulant maximization [Common, 1994]
  - Infomax [Bell & Sejnowski, 1995]
  - Likelihood maximization [Pham 1992, Pearlmutter & Parra 1996]
  - Fixed point ICA [Hyvarinen & Oja, 1997]

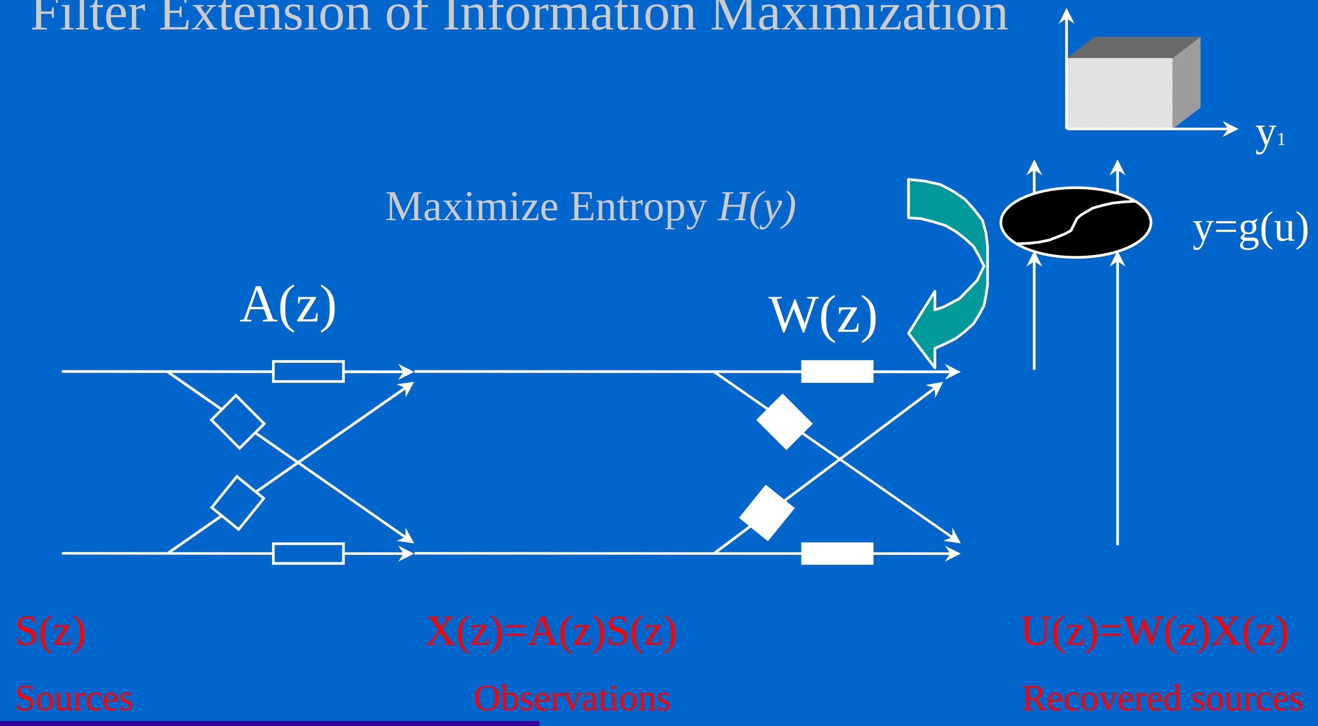
- **Challenges:**
  - Noisy Observations (nonstationary noise)
  - Non square mixing (overcomplete and undercomplete models)
  - Nonlinear mixing
  - Dependent sources

Multichannel Blind Deconvolution

$$\mathbf{x}(t) \;=\; \sum_{\tau=0}^{P} \mathbf{A}(\tau)\,\mathbf{s}(t-\tau) \;+\; \mathbf{n}(t)$$

Yellin & Weinstein, 1996
Torkkola, 1997
Lambert, 1997
Douglas, 1999
…

Filter Extension of Information Maximization

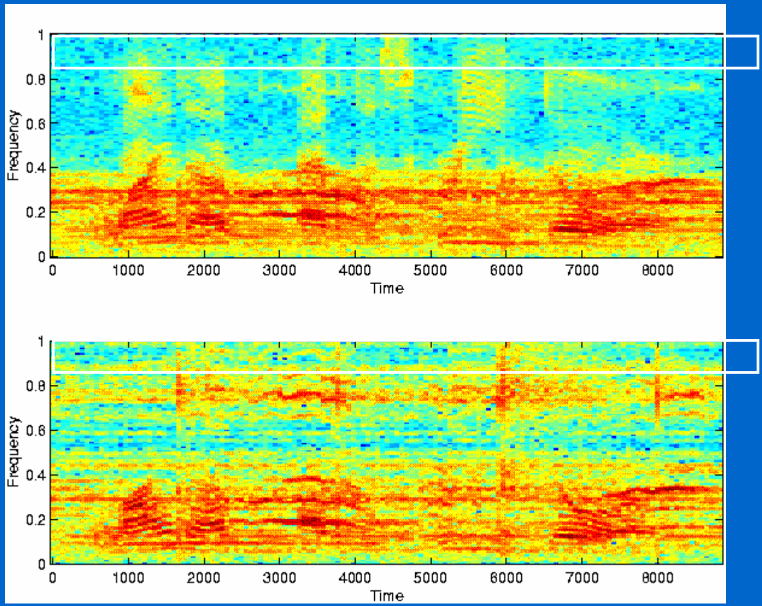


Time-Delayed Decorrelation

-Solving the eigenvalue problem
(Tong et al., 1992; Belouchrani et al. 1993, Molgedey & Schuster, 1994):
Joint diagonalization of the covariance matrix and time-delayed cov.-matrix.

$$\mathbf{C}_0 = \left\langle \mathbf{x}(t)\mathbf{x}(t)^T \right\rangle \qquad \mathbf{C}_\tau = \left\langle \mathbf{x}(t)\mathbf{x}(t-\tau)^T \right\rangle \qquad \left[\mathbf{C}_0\mathbf{C}_\tau^{-1}\right]\mathbf{A} = \mathbf{A}\left[\Lambda_0\Lambda_\tau^{-1}\right]$$

- Extension to filters (Ehlers & Schuster, 1999):
Solve eigenvalue problem for each frequency bin.
Separation quality depends on the spectral overlap in the data.



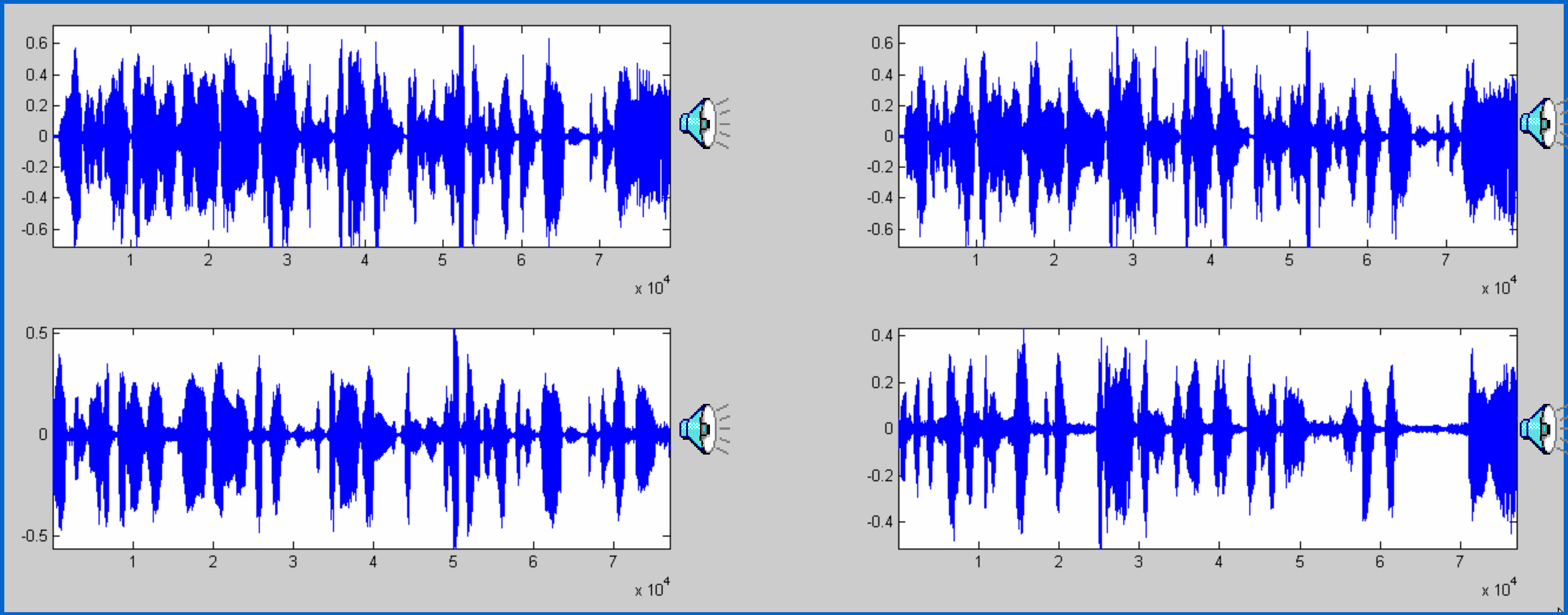
TDD $\mathbf{C}_f(z)$

$$\left[\mathbf{C}_0(z)\mathbf{C}_\tau^{-1}(z)\right]\mathbf{A}(z) = \mathbf{A}(z)\left[\Lambda_0(z)\Lambda_\tau^{-1}(z)\right]$$

Problem of scaling and permutation:
Possible solutions:
- Murata et al. (1999): back-projection, cross-correlation
- Parra (2000): Sinc filtering
- Annemueller (2001)

Demonstration

-Live BSS demonstration at NSF workshop
-Stereo recording of Al Bregman and Te-Won Lee speaking simultaneously
-Data is 10 seconds long, 16kHz sampling frequency



Issues in Multi Channel Speech Separation

- Sources localization
  - Directivity information
  - Singular mixing conditions
- Mismatch between sensor and source number
- Tracking non-stationary sources
  - Adaptation time versus convergence time
  - Source identification
- Incorporation of speech models
  - HMM trained models
  - Speech cues, frequency grouping

New approaches for solving those issues tackled by several groups: Makino, Sawada, Rosca, Ziehe

Single Microphone Source Separation

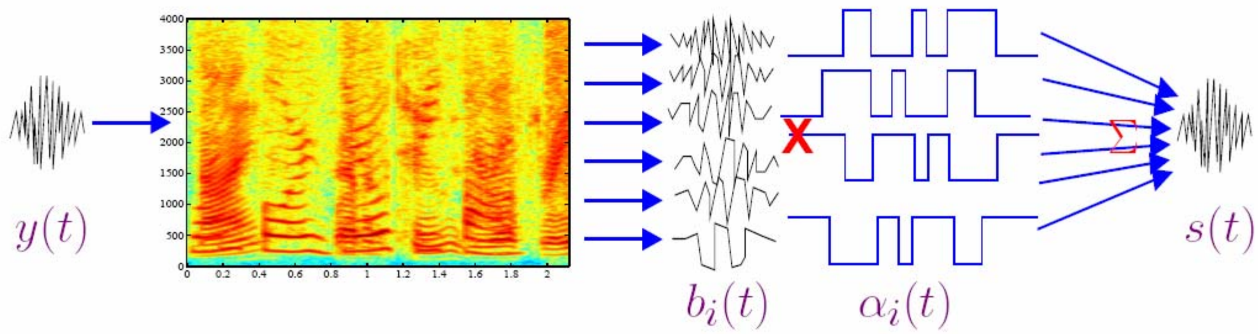
- Roweis, 2000, 2003
  - Sparse speech code for spectrogram mixtures
  - Refiltering (masking) approach
- Jang & Lee, 2002
  - Learning basis functions resulting in sparse speech codes
  - Soft masking approach
- Pearlmutter et al., 2004
  - Learning highly overcomplete feature set, separation by sparse decomposition

Roweis, 2000, 2003

Masking (Refiltering) Paradigm

Non-constant reweighting of original multiband signals $b_i(t)$.

$$\underbrace{s(t)}_{\text{est. source}} = \overbrace{\alpha_1(t)}^{\text{mask 1}}\,\underbrace{b_1(t)}_{\text{band 1}} + \overbrace{\alpha_2(t)}^{\text{mask 2}}\,\underbrace{b_2(t)}_{\text{band 2}} + \ldots + \overbrace{\alpha_K(t)}^{\text{mask K}}\,\underbrace{b_K(t)}_{\text{band K}}$$

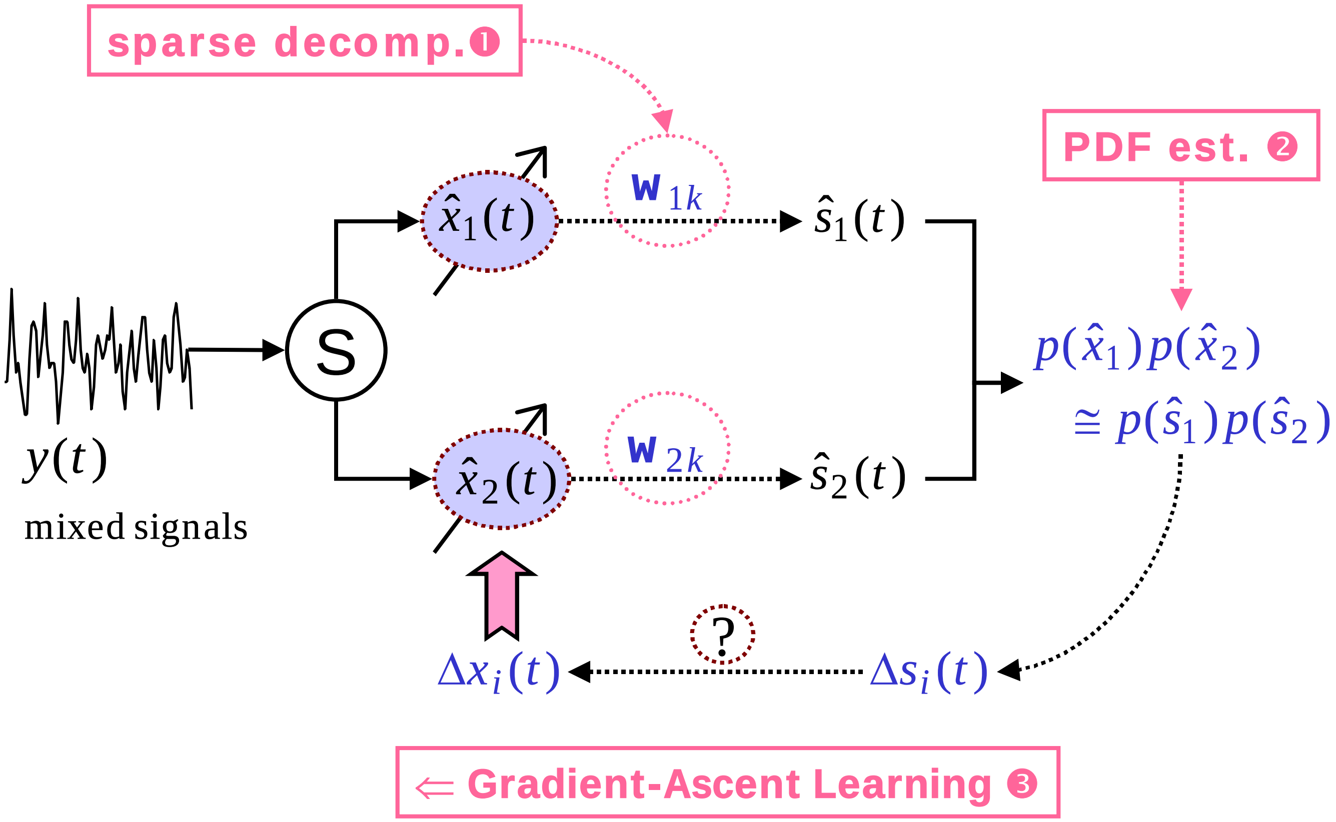


$\alpha_i(t)$ are gain knobs on each subband which we can twist over time to bring bands in and out of the recovered source as needed.

Works extremely well if the masking signals are chosen well.

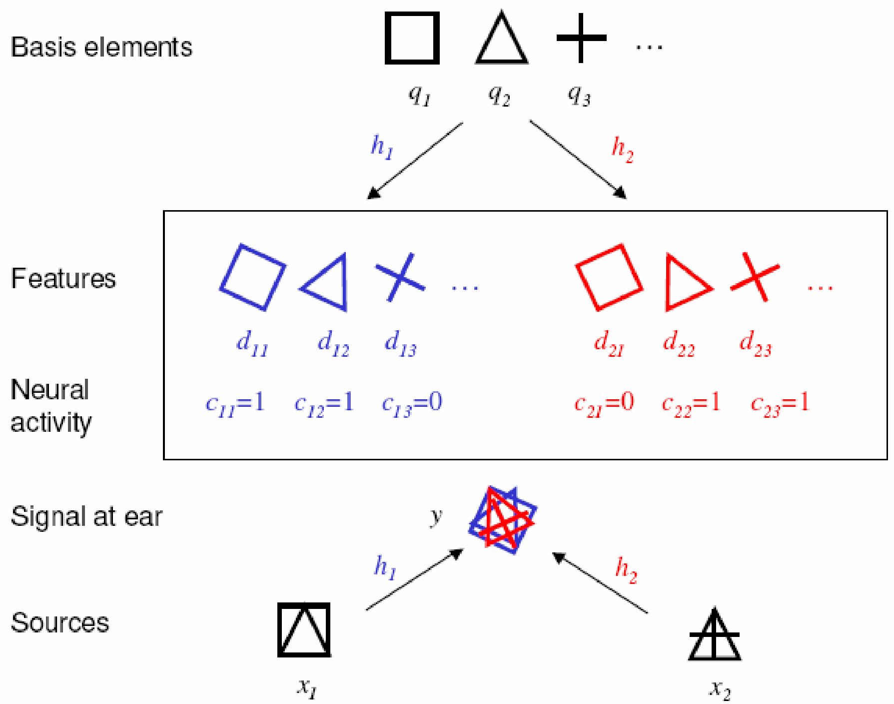
Jang & Lee, 2002

Proposed Method



Pearlmutter et al., 2004

Cartoon of Algorithm



Shapes represent acoustic features and colours/tilts represent HRTF filtering. Separation accomplished by sparse decomposition using basis of HRTFs $*$ source dictionary.

Future Research Directions

- Solve binaural acoustic source separation for underdetermined case and make use of learned features

$$\mathbf{x}(t) = \sum_{\tau=0}^{P} \mathbf{A}(\tau)\, \mathbf{s}(t-\tau) + \mathbf{n}(t)$$

2 microphones   Acoustic mixing   Overcomplete   Sparse representation