Blind sparse-nonnegative (BSN) channel identification for acoustic TDOA estimation

Yuanqing Lin (1) *Jingdong Chen* (2) *Youngmoo Kim* (3) *Daniel D. Lee* (1)

(1) GRASP Laboratory, Department of Electrical and Systems Engineering, University of Pennsylvania

(2) Bell Labs, Alcatel-Lucent

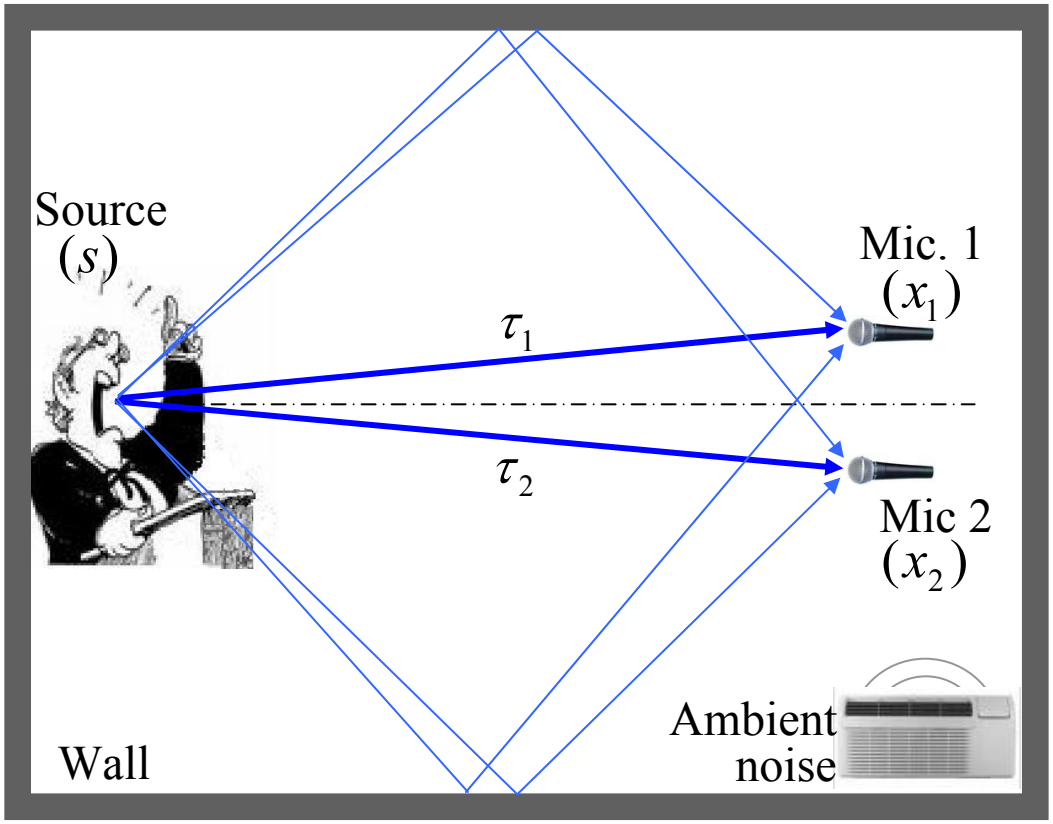
(3) Electrical and Computer Engineering Department, Drexel University

WASPAA 2007

Outline

1. Introduction
   - Problem of TDOA estimation
   - Existing methods
   - Our new method
2. Our proposal: blind sparse-nonnegative channel identification
   - Room acoustic model
   - Convex formulation
   - Bayesian framework (for inferring optimally sparse solutions)
3. Results
   - Simulations
   - Experiments in real acoustic environments

TDOA estimation -- problem description



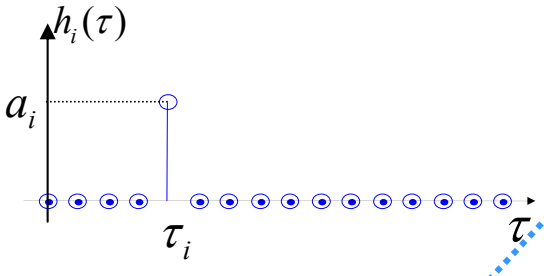
- Goal of TDOA estimation
  - Compute $\Delta t = \tau_2 - \tau_1$
  - given only microphone signals
- Challenge:
  - Ambient noise
  - Echoes (reverberation)

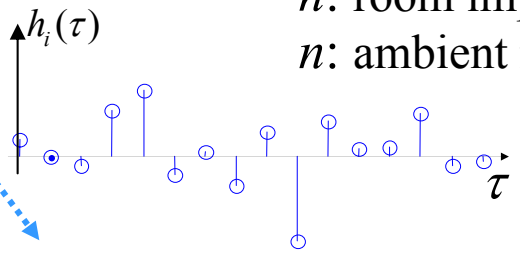
- Open problem: robustly estimating TDOA in noisy and reverberant environments.

The existing methods -- blind channel identification

$$x_i(k) = s(k) * h_i + n_i(k) \quad i = 1,2$$

x: microphone signal
s: source signal
h: room impulse response
n: ambient noise





Cross-correlation:

$$a_1 = 1, \quad \tau_1 = 0$$

$$c(\tau) = \frac{\sum_k x_1(k-\tau)x_2(k)}{\sum_k x_1^2(k)}$$

$$\tau_2 = \arg\max_{dt} c(\tau)$$

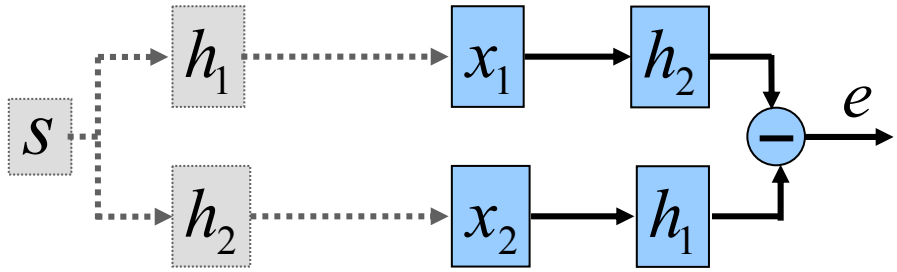
$$a_2 = c(\tau_2)$$

Eigenvalue decomposition approach

[J. Benesty, 2000]

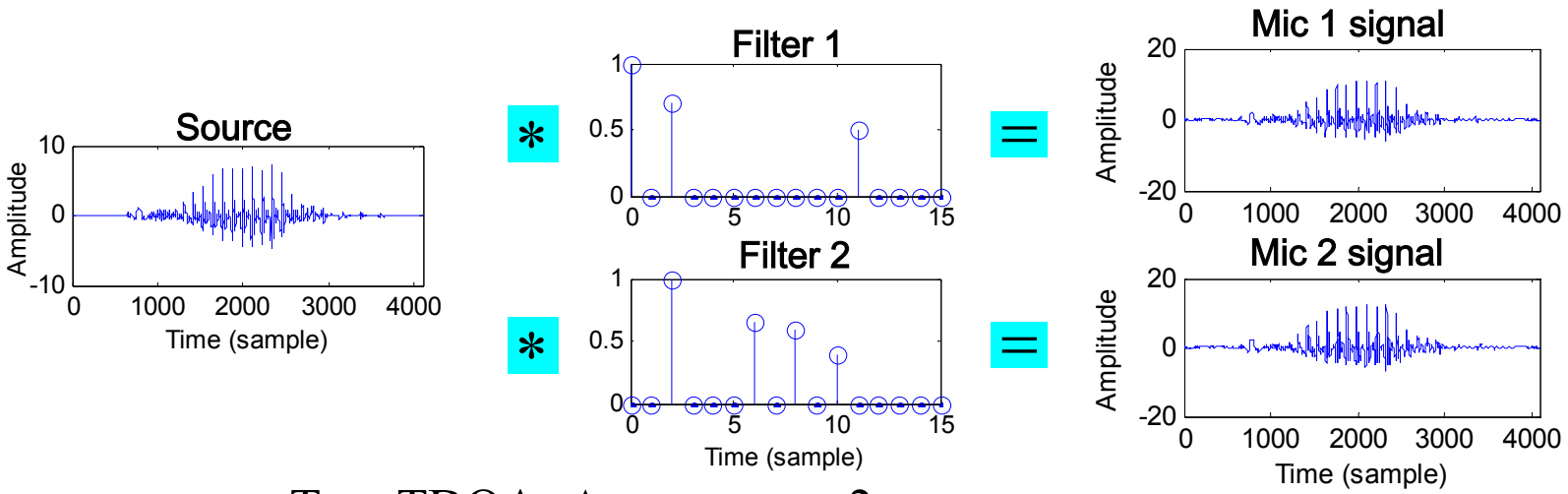
$$h_1^*, h_2^* = \arg\min_{h_1,h_2} \frac{1}{2}\|x_2 * h_1 - x_1 * h_2\|^2$$

$$\text{S.T.} \quad \|h_1\|^2 + \|h_2\|^2 = 1$$



Noiseless: $x_2 * h_1 = s * h_2 * h_1 = s * h_1 * h_2 = x_1 * h_2$

A simulated example

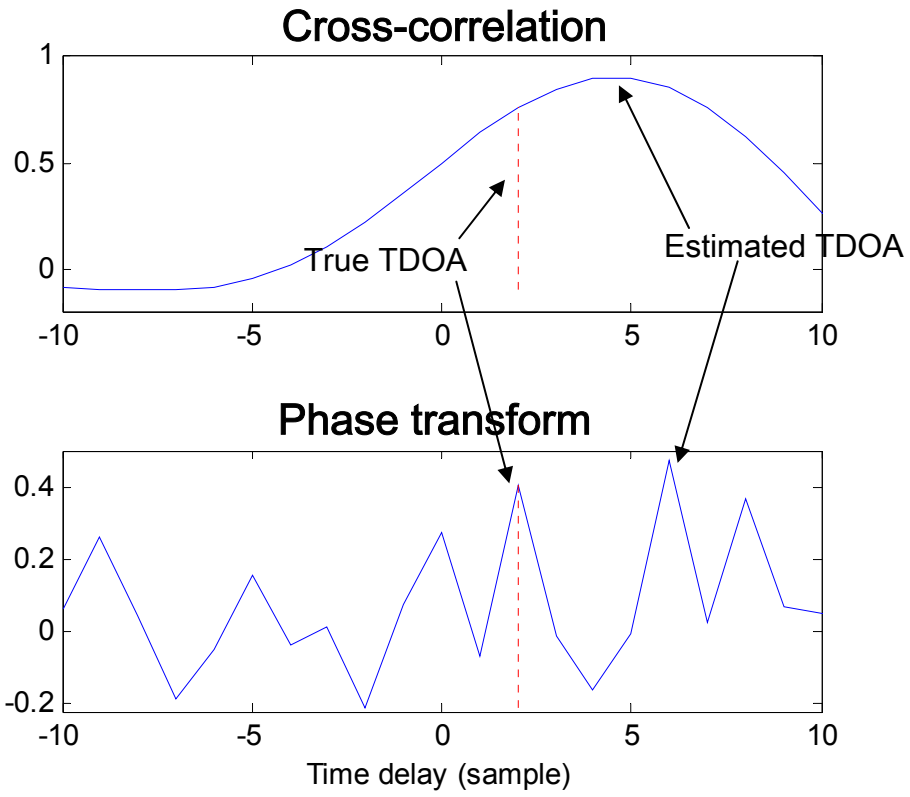


$$\text{True TDOA}: \Delta t = \tau_2 - \tau_1 = 2$$

- TDOA estimation: given only the *reverberant* microphone observations.

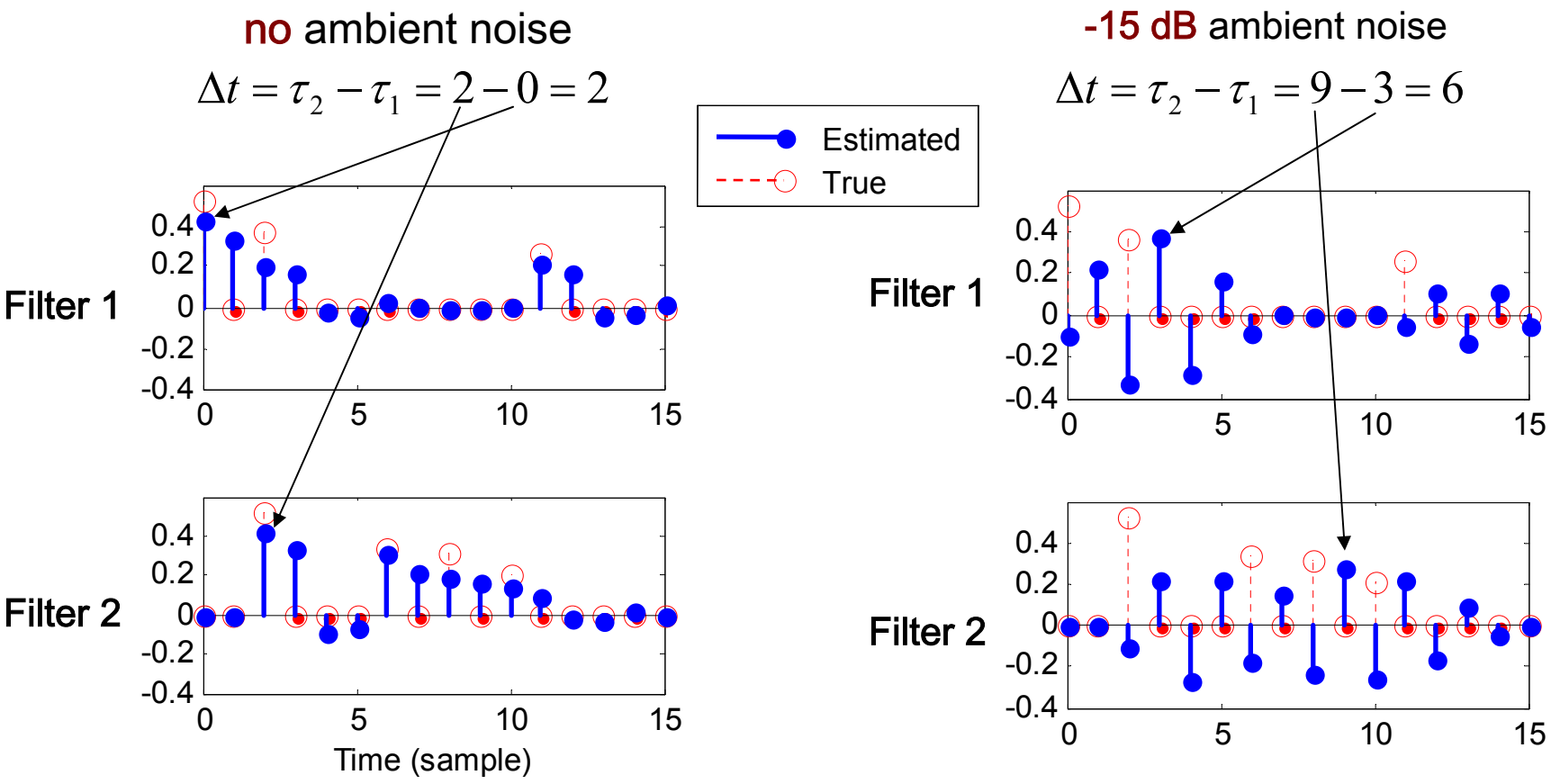
Generalized cross-correlation

Problematic…



☹ Generalized cross-correlation: sensitive to reverberant echoes.

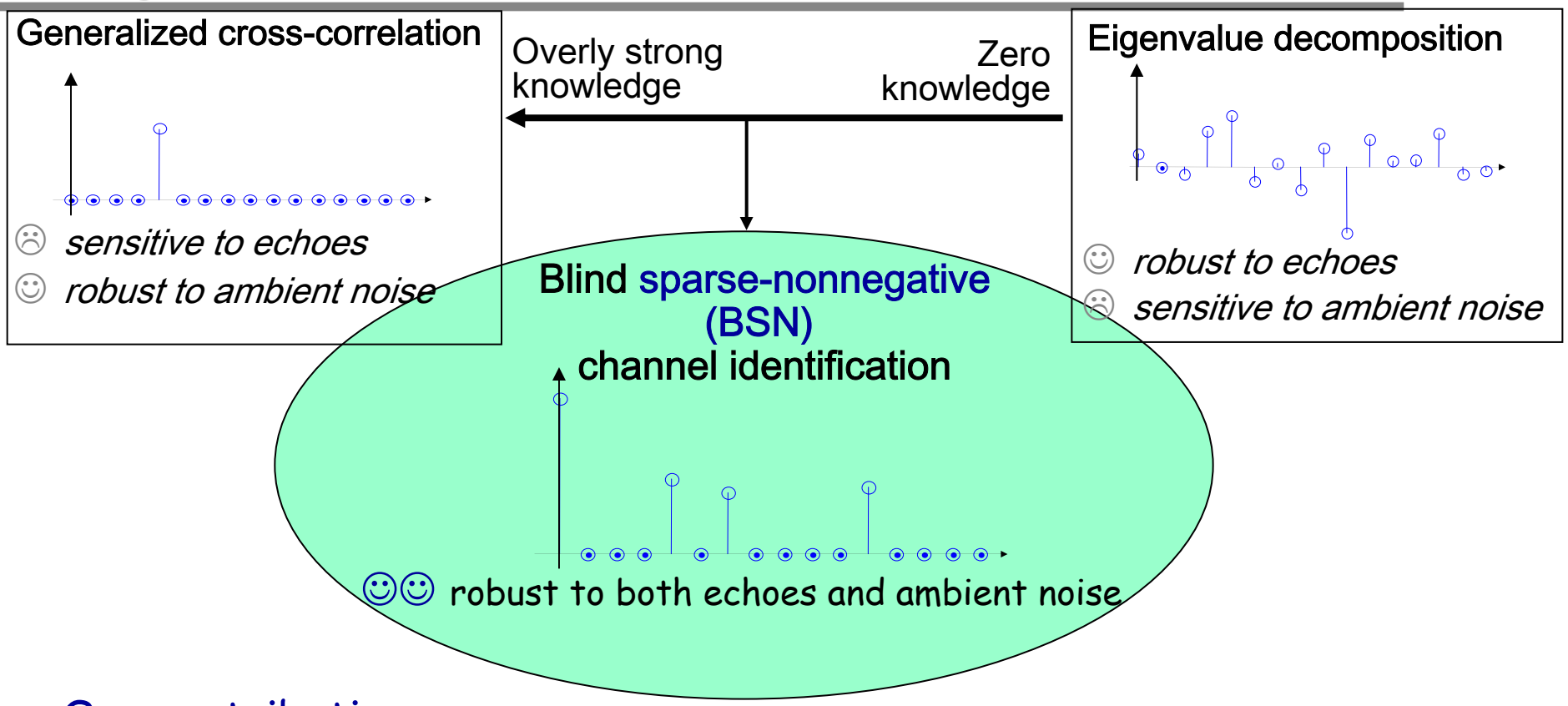
Eigenvalue decomposition approach



☺ effective in reverberant environments

☹ sensitive to ambient noise

Why our new method?



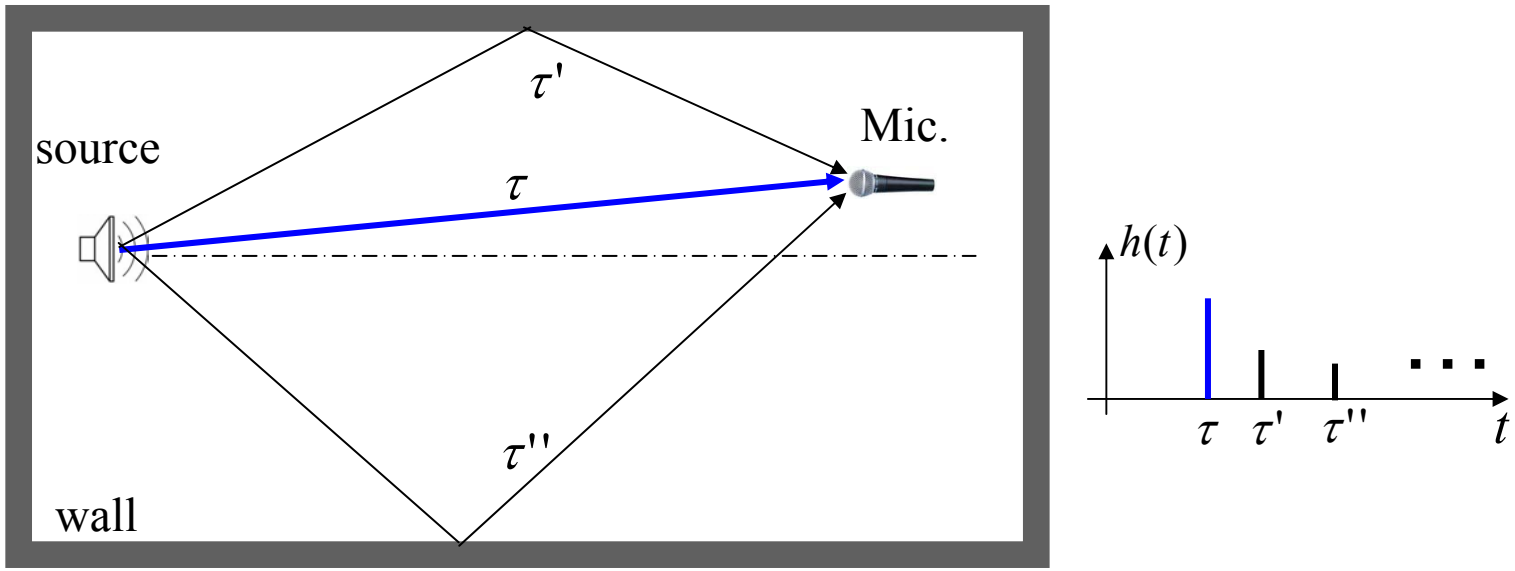
- Our contributions:

  1) sparse-nonnegative room acoustic model;
  2) convex formulation;
  3) Bayesian framework (for inferring optimally sparse solution).

Outline

1. Introduction
   - Problem of TDOA estimation
   - Existing methods
   - Our new method
2. Our proposal: blind sparse-nonnegative channel identification
   - Room acoustic model
   - Convex formulation
   - Bayesian framework (for inferring optimally sparse solutions)
3. Results
   - Simulations
   - Experiments in real acoustic environments

Acoustic model -- room impulse response



- **Image Model:** the FIR filter modeling a room impulse response is nonnegative and sparse. [J. Allen *et. al*, 1979]

Enforcing sparse and nonnegative priors?

$$\mathbf{h}_1^*, \mathbf{h}_2^* = \arg\min_{\mathbf{h}_1, \mathbf{h}_2} \frac{1}{2} \left\| \mathbf{X}_2 \mathbf{h}_1 - \mathbf{X}_1 \mathbf{h}_2 \right\|^2 \qquad \mathbf{X}_i : \text{convolution matrix}$$

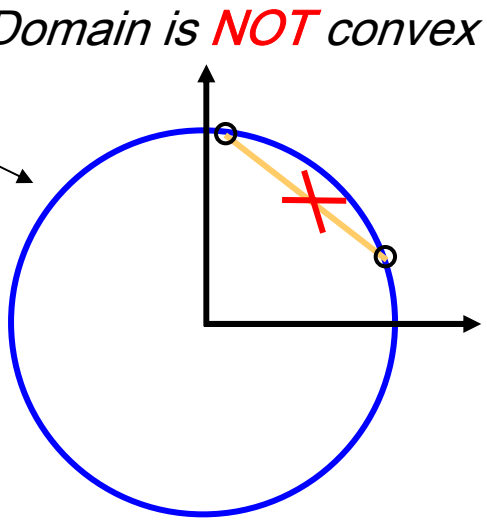
S.T.

$$\left\| \mathbf{h}_1 \right\|^2 + \left\| \mathbf{h}_2 \right\|^2 = 1$$

$$\mathbf{h}_1 \geq 0$$

$$\mathbf{h}_2 \geq 0$$

$\mathbf{h}_1$ and $\mathbf{h}_2$ are sparse



☹ Optimization is not convex even with only nonnegative constraints.
-- extremely hard to solve…

Convex formulation

$$\mathbf{h}_1^*, \mathbf{h}_2^* = \arg\min_{\mathbf{h}_1, \mathbf{h}_2} \frac{1}{2} \left\| \mathbf{X}_2 \mathbf{h}_1 - \mathbf{X}_1 \mathbf{h}_2 \right\|^2$$

S.T. ~~$\|\mathbf{h}_1\|^2 + \|\mathbf{h}_2\|^2 = 1$~~

Convex constraint → $h_1(0) = 1$

- Equivalent when noiseless;
- Align $h_1^*(0)$ with the largest coefficient in h_1

  -- remove scaling and phase degeneracy.

☺ Optimization is convex, indeed an ordinary least-squares problem;

☺ In fact, more robust to ambient noise.

Blind sparse-nonnegative (BSN) channel identification

$$\mathbf{h}_1^*, \mathbf{h}_2^* = \arg\min_{\mathbf{h}_1, \mathbf{h}_2} \left\| \mathbf{X}_2\mathbf{h}_1 - \mathbf{X}_1\mathbf{h}_2 \right\|^2 + \lambda'\left(|\mathbf{h}_1| + |\mathbf{h}_2|\right)$$

S.T. $h_1(0) = 1$

$\mathbf{h}_1 \geq 0$

$\mathbf{h}_2 \geq 0$

Enforcing sparsity by l_1-norm regularization [S. S. Chen *et. al*, 1998]

Enforcing nonnegativity by nonnegative constraints

☺ Optimization is convex

--- easy to solve with guaranteed global convergence.

Determining the optimal regularization parameter

$$\mathbf{h}_1^*, \mathbf{h}_2^* = \arg\min_{\mathbf{h}_1, \mathbf{h}_2} \left\| \mathbf{X}_2\mathbf{h}_1 - \mathbf{X}_1\mathbf{h}_2 \right\|^2 + \lambda' \left(\left| \mathbf{h}_1 \right| + \left| \mathbf{h}_2 \right| \right)$$

$$\text{S.T.} \quad h_1(0) = 1, \qquad \mathbf{h}_1 \geq 0, \qquad \mathbf{h}_2 \geq 0$$

Bayesian framework:

$$\lambda' = \frac{2L\sigma_0^2 \left\langle \left\| \mathbf{h}_1 \right\|^2 + \left\| \mathbf{h}_2 \right\|^2 \right\rangle}{\left\langle \left| \mathbf{h}_1 \right| + \left| \mathbf{h}_2 \right| \right\rangle}$$

σ_0^2 : ambient noise level

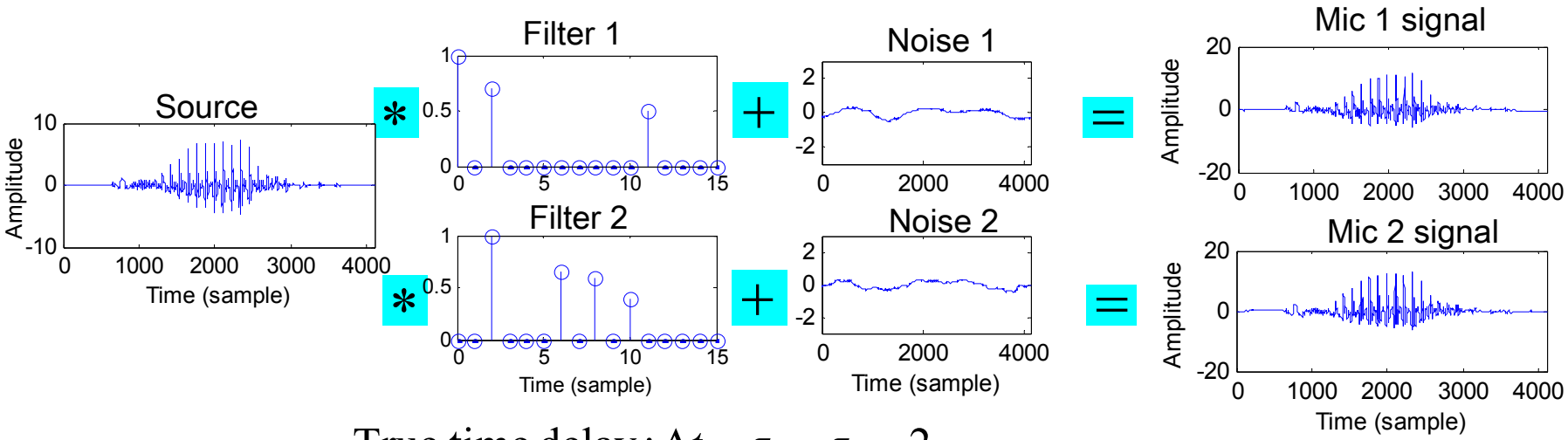
L : filter length

- The filter statistics: inferred in a Bayesian framework
  or computed from prior knowledge (e.g. reverberation time).

Outline

1. Introduction
   - Problem of TDOA estimation
   - Existing methods
   - Our new method
2. Our proposal: blind sparse-nonnegative channel identification
   - Room acoustic model
   - Convex formulation
   - Bayesian framework (for inferring optimally sparse solutions)
3. Results
   - Simulations
   - Experiments in real acoustic environments

The simulated example

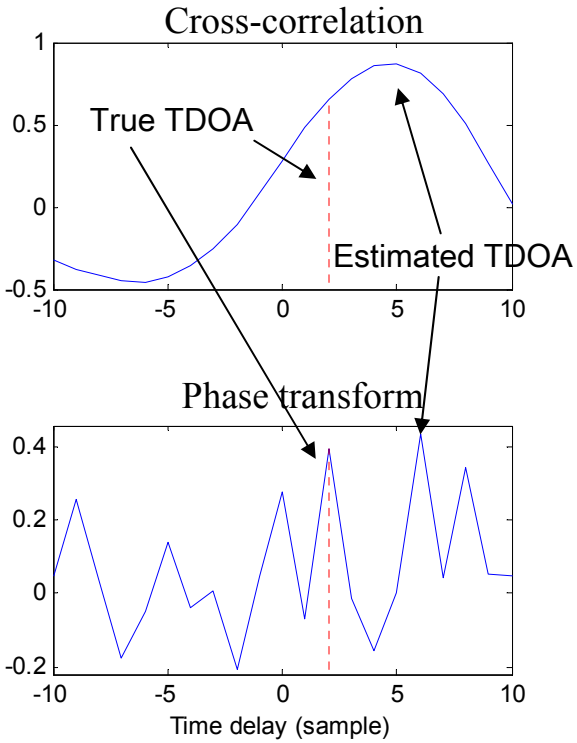


$$\text{True time delay}: \Delta t = \tau_2 - \tau_1 = 2$$

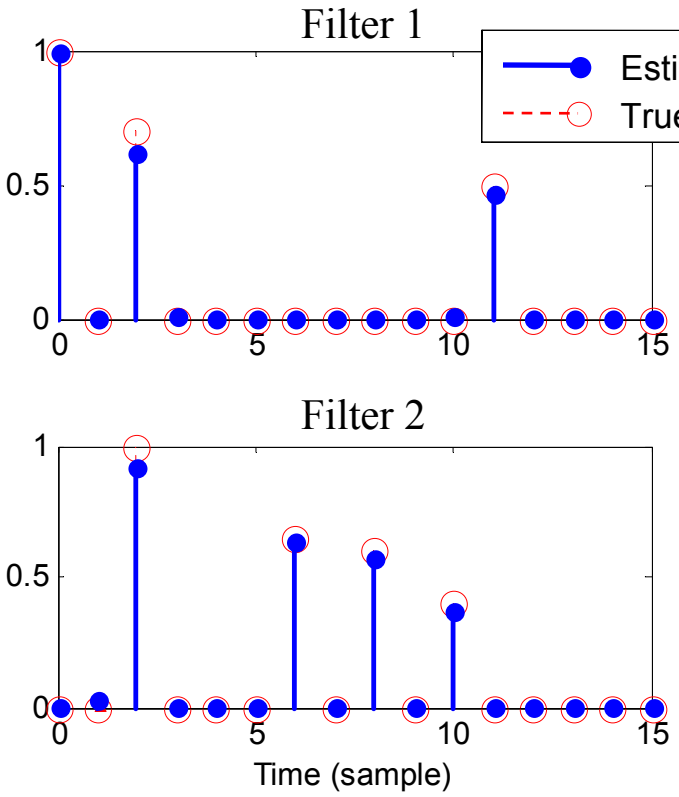
- TDOA estimation: given only the *reverberant* microphone observations.

Simulations -- results

Generalized cross-correlation

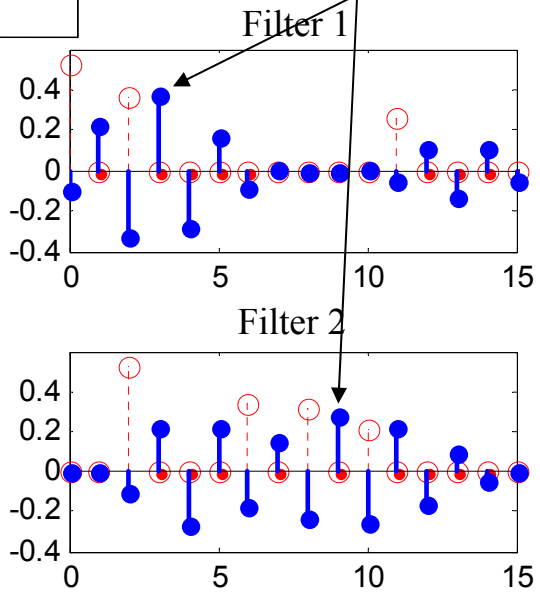


Blind sparse-nonnegative (BSN) channel identification



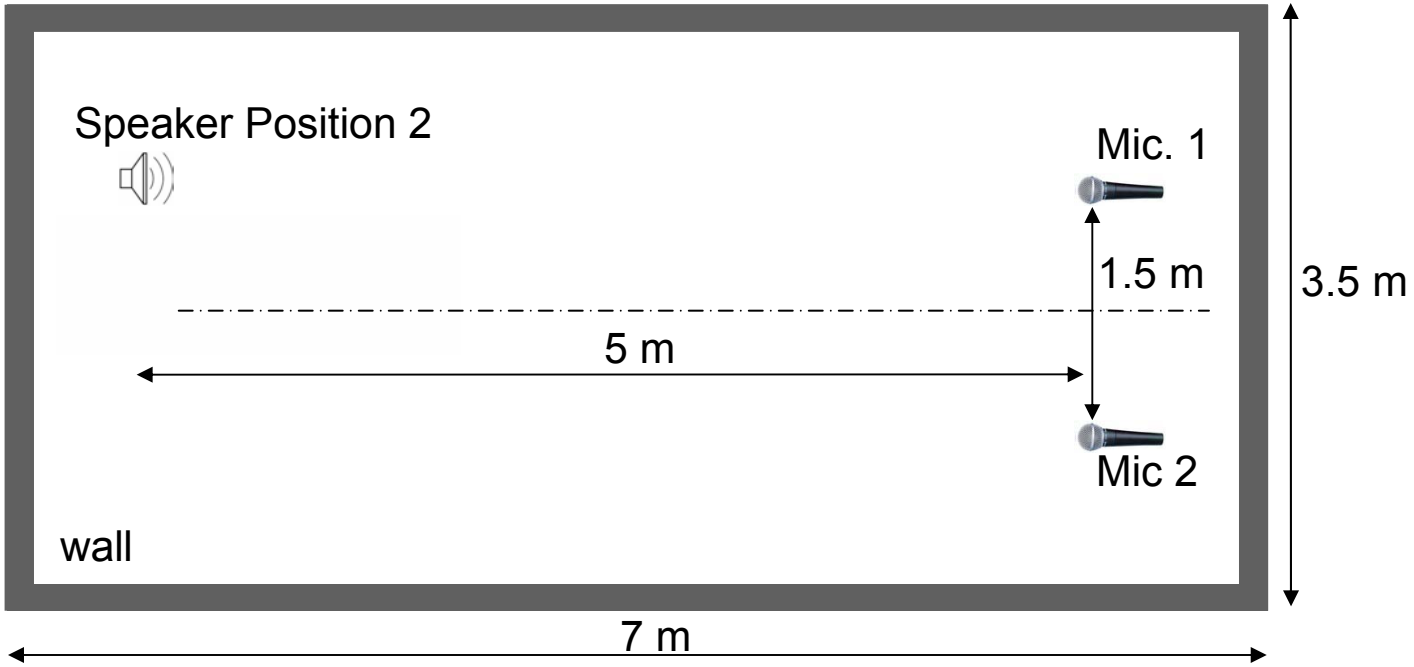
Eigenvalue decomposition

$\Delta t = \tau_2 - \tau_1 = 9 - 3 = 6$



- Blind sparse-nonnegative (BSN) blind channel identification:
  -- yields accurate filter estimates, so as TDOA estimates.

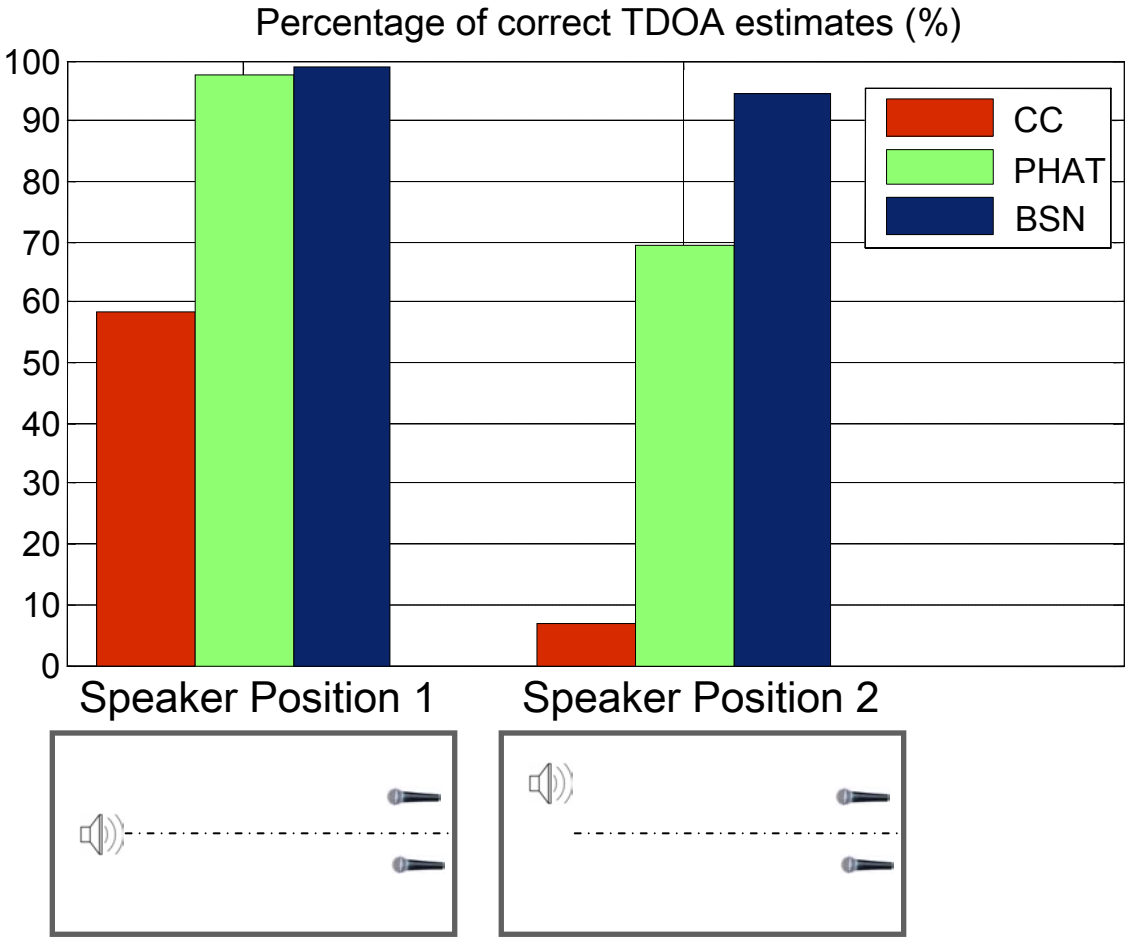
Experiments -- setup



Playback speech source: 100 sentences ; **Sampling rate:** 16,000 Hz;
Window size: 4096 samples; **Filter length:** 2048 samples.

- Speaker Position 2: Mic 1 picked up some strong reflections.

Experiments -- results



- Our method (BSN): performed consistently better than conventional methods.
- Eigenvalue decomposition: too ill-conditioned to yield meaningful estimates
  -- filter length 2048, data length 4096.

Summary

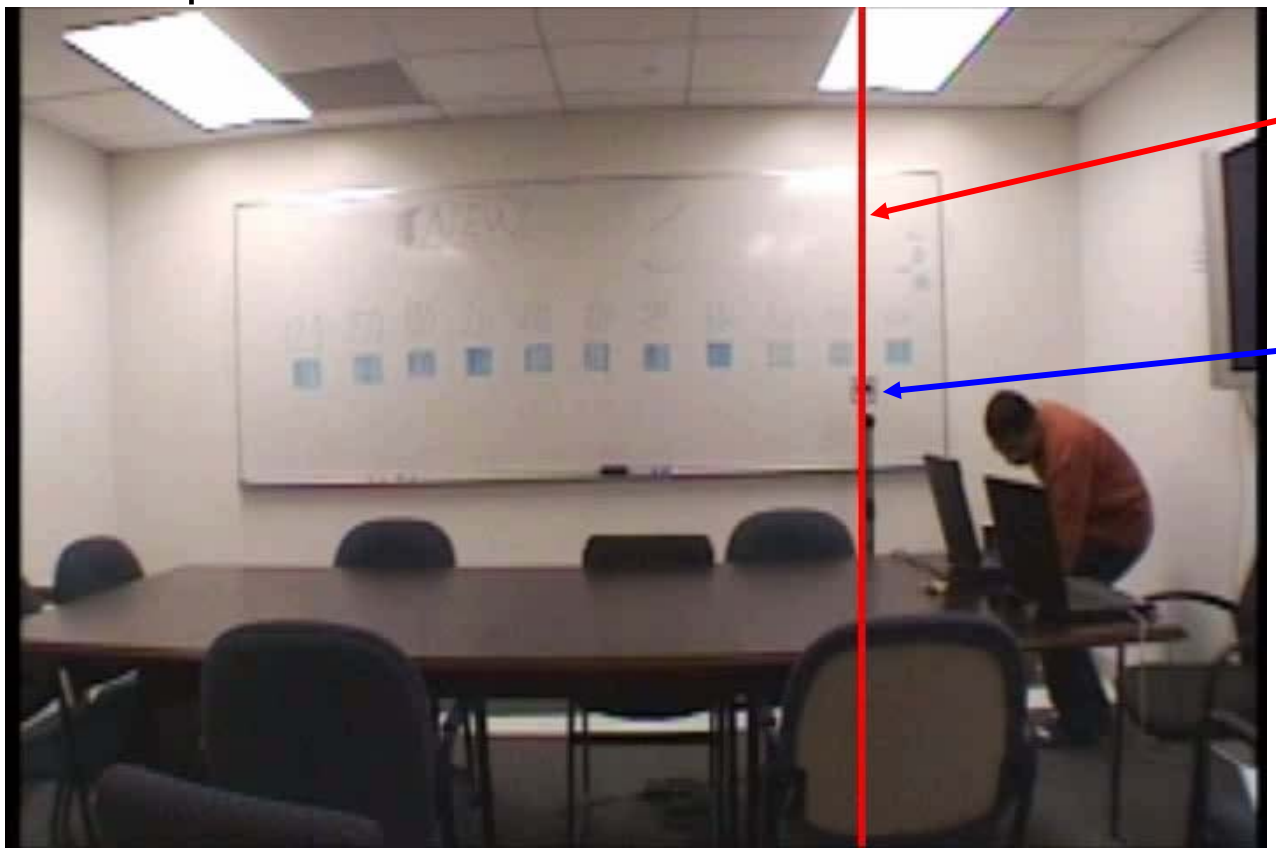
- Blind sparse-nonnegative (BSN) channel identification (3 key components):

  1) room acoustic model: sparse and nonnegative;
  2) convex formulation;
  3) Bayesian framework (for inferring optimally sparse solution).

- BSN channel identification is robust to both echoes and ambient noise.

More in demo session tonight…

- More examples of TDOA estimation in real acoustic environments.



- Applications to *speech dereverberation in real acoustic environments*.
  -- preview of our NIPS paper