Blind channel identification for speech dereverberation using l_1-norm sparse learning

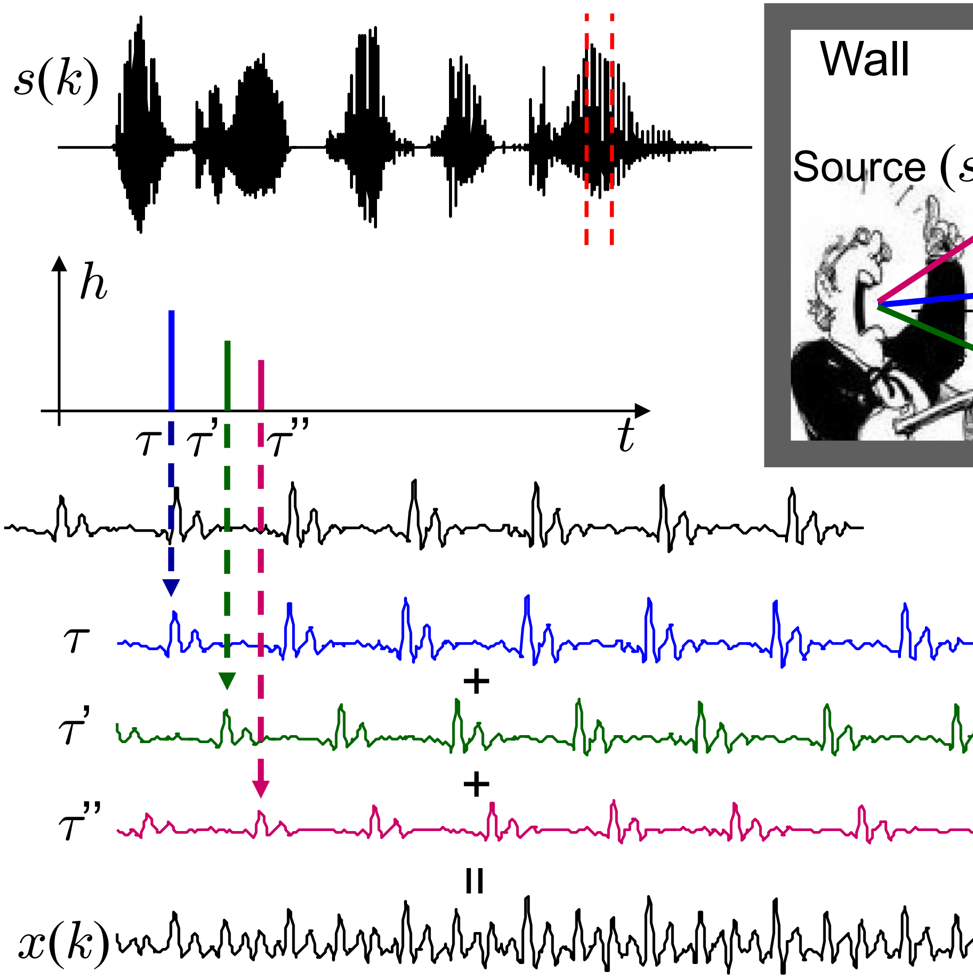
Yuanqing Lin (1) *Jingdong Chen* (2) *Youngmoo Kim* (3) *Daniel D. Lee* (1)

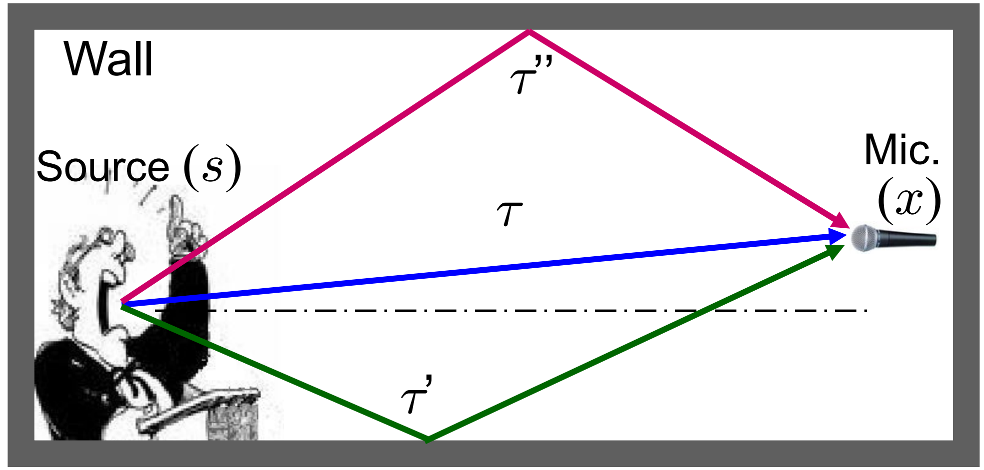
(1) GRASP Laboratory, Department of Electrical and Systems Engineering, University of Pennsylvania

(2) Bell Labs, Alcatel-Lucent

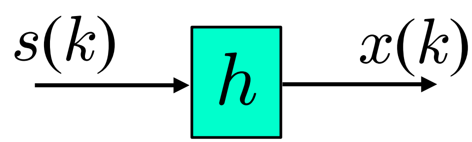
(3) Electrical and Computer Engineering Department, Drexel University

Room reverberation





Model (convolution):

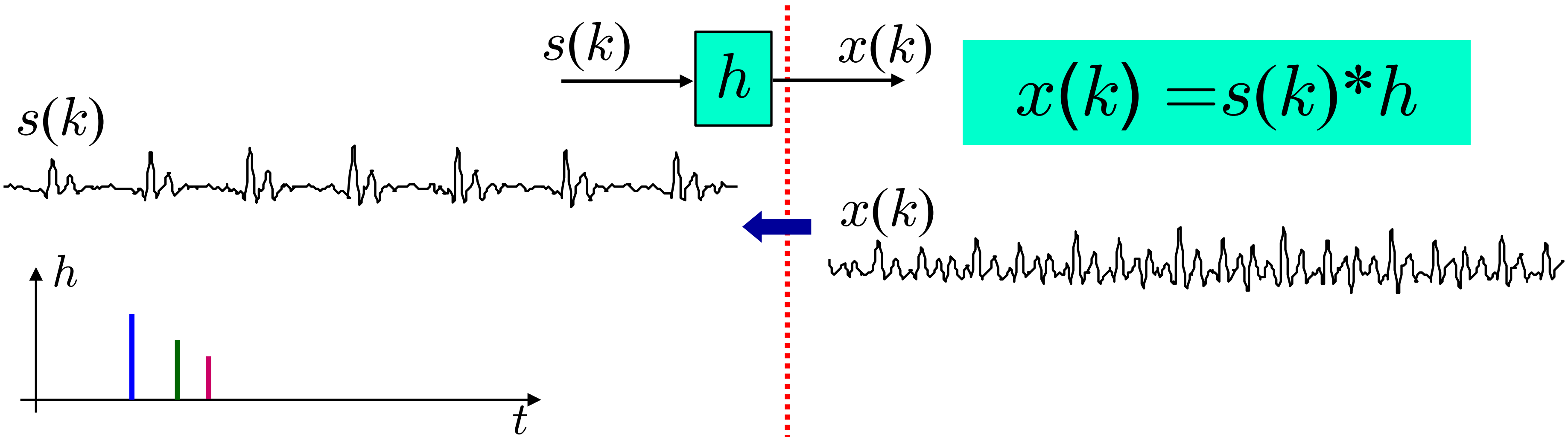


$$x(k) = s(k) * h$$

◆ Reverberation (convolution): temporal mixing or blurring.

Speech dereverberation

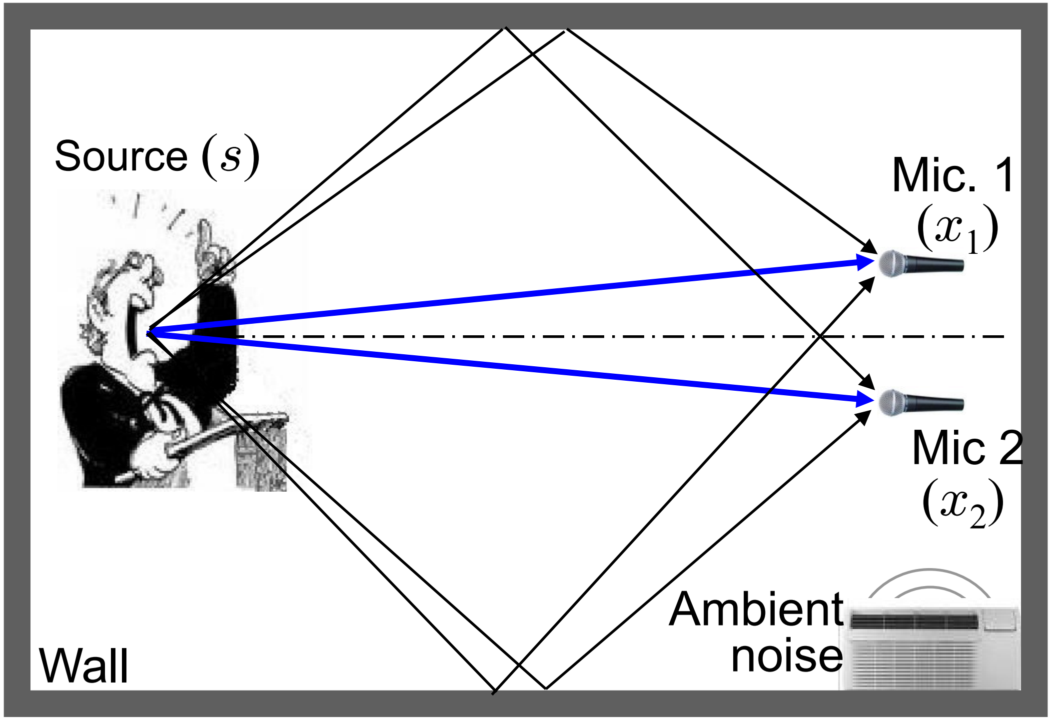
Blind channel identification:



$$x(k) = s(k) * h$$

- Dereverberation (deconvolution): temporal unmixing or deblurring.

Blind channel identification
-- problem description



- **Model:**

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

x_i : microphone signals
s : source signal
h_i : channel filters
n_i : ambient noise

- **Goal:**

Compute h_1 and h_2 given **only** x_1 and x_2.

- **Challenge:**

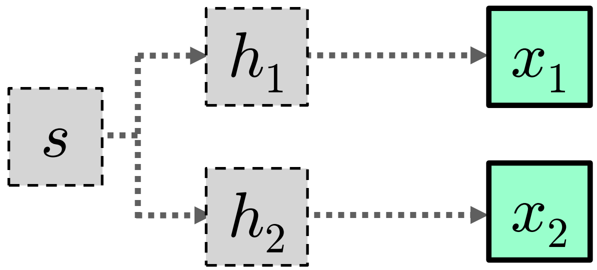
Solution degeneracy (missing source):

$$x_i(k) = [s(k)*a^{-1}]*[a*h_i] + n_i(k)$$

a: arbitrary invertible filter

◆ Open problem: blind channel identification in real acoustic environments.

Existing methods (1) -- source statistics



Prior: $P(s)$

$$h_1^*, h_2^* = \arg\max_{h_1,h_2} \int_s P(x_1, x_2 | s, h_1, h_2) P(s) ds$$

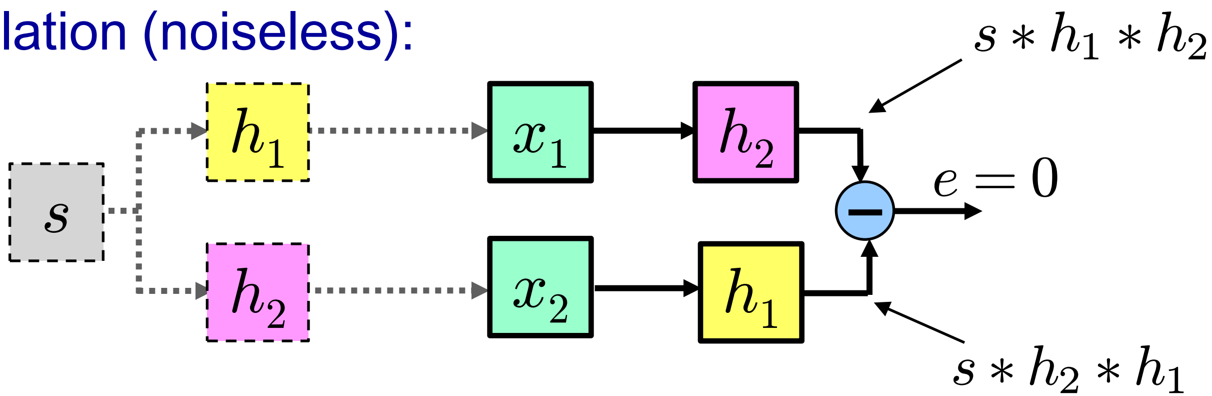
(A. J. Bell *et al.*, 1995; D. J. C. MacKay, 1999; A. Hyvarinen *et al.*, 2001; H. Attias *et al.*, 2001)

☹ Difficulties in assuming source statistics:

- Need long data (hard to obtain in time-varying real acoustic environments)
- Whitening side effect
- Non-stationarity of speech

Existing methods (2) -- cross-relation

Cross-relation (noiseless):



- Convolution operation commutes.

Eigenvalue decomposition approach:

$$h_1^*, h_2^* = \arg\min_{h_1, h_2} \frac{1}{2} \|x_2 * h_1 - x_1 * h_2\|_2^2$$

$$\text{S.T.} \quad \|h_1\|_2^2 + \|h_2\|_2^2 = 1$$

(L. Tong *et al.*, 1994; Y. Huang *et al.*, 2005)

☺ Free of difficulties in assuming source priors

☹ Very sensitive to ambient noise

Our method -- blind ***sparse*** channel identification

Methods \ Prior Knowledge	source	channels
Source priors	Yes	No
Eigen-decomposition	No	No
BSCI (Our method)	**No**	**Yes**

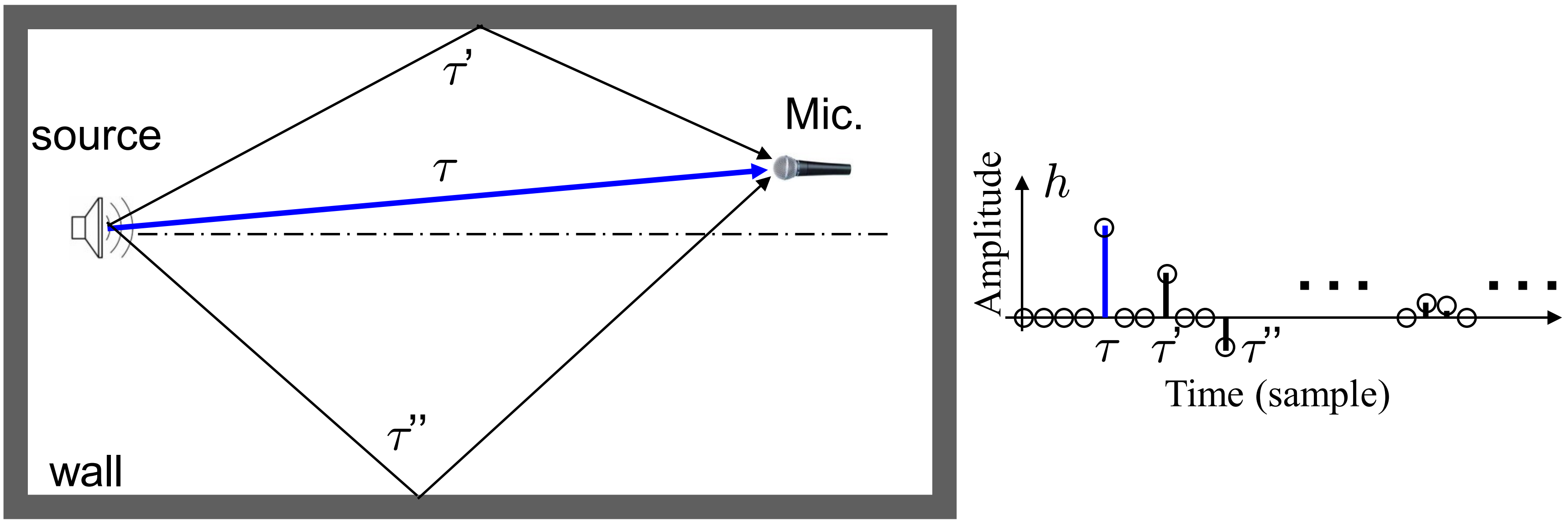
- What our method wants to achieve:
  - Free of difficulties in assuming source statistics
  - Robust to ambient noise

Our contributions:

1) Sparse acoustic room impulse response model;
2) Convex formulation;
3) l_1-norm sparse Bayesian learning

-- for inferring the optimal sparsity regularization parameters.

Acoustic model -- room impulse response



- Acoustic room impulse responses:
  --- modeled by sparse finite-impulse-response (FIR) filters.

(J. Allen *et al.*, 1979; D. L. Duttweiler, 2000)

Enforcing sparsity prior

$$\mathbf{h}_1^*, \mathbf{h}_2^* = \arg\min_{\mathbf{h}_1, \mathbf{h}_2} \frac{1}{2} \|\mathbf{X}_2 \mathbf{h}_1 - \mathbf{X}_1 \mathbf{h}_2\|_2^2 + \lambda' (|\mathbf{h_1}| + |\mathbf{h_2}|)$$

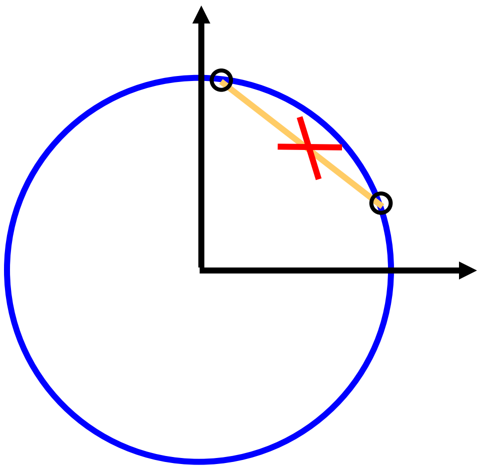
$$\text{S.T.} \quad \|\mathbf{h}_1\|_2^2 + \|\mathbf{h}_2\|_2^2 = 1$$

Enforcing sparsity by l_1-norm regularization

(R. Tibshirani, 1996;
B.A. Olshausen *et al.*, 1996;
S. S. Chen *et al.*, 1998)

$\mathbf{X}_i$: convolution matrix

Domain is NOT convex



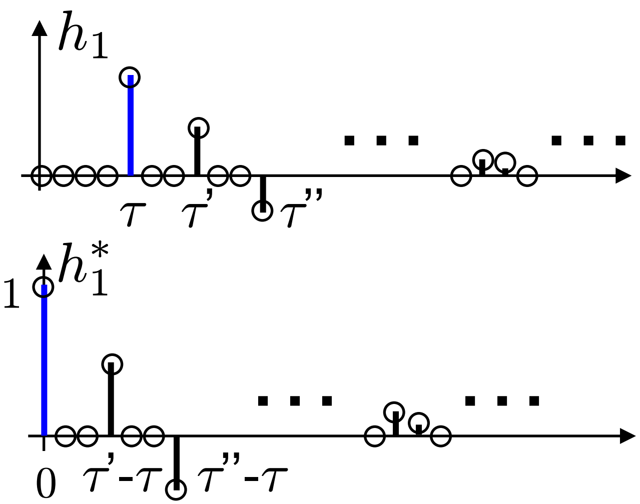
☹ Optimization is very hard to solve…

Convex formulation

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

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

Convex constraint → $h_1(0) = 1$



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

-- remove both phase and scaling degeneracy.

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

☺ OLS: more robust to ambient noise.

Blind sparse channel identification (BSCI)

$$\mathbf{h}_1^*, \mathbf{h}_2^* = \arg\min_{\mathbf{h}_1,\mathbf{h}_2} \frac{1}{2}\|\mathbf{X}_2\mathbf{h}_1 - \mathbf{X}_1\mathbf{h}_2\|_2^2 + \lambda'(|\mathbf{h_1}| + |\mathbf{h_2}|)$$

$$\text{S.T.} \quad h_1(0) = 1$$

Enforcing sparsity by l_1-norm regularization

☺ Optimization (Lasso) is convex

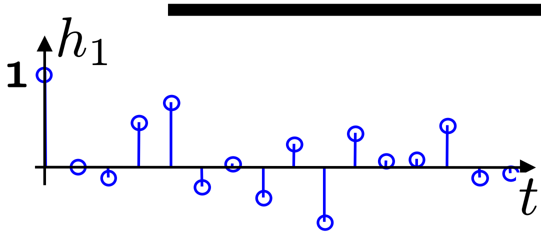
--- easy to solve with guaranteed global convergence.

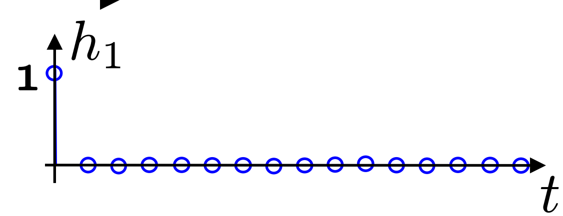
(S. J. Wright, 1997; B. Efron *et al.*, 2003; Y. Lin *et al.*, 2006; S. J. Kim *et al.*, 2007)

◆ Sparsity regularization parameter λ'

--- critical for deriving appropriate solutions.

$\lambda' = 0$ → $\lambda' = +\infty$





How to determine the regularization parameter

$$\mathbf{h}_1^*, \mathbf{h}_2^* = \arg\min_{\mathbf{h}_1, \mathbf{h}_2} \frac{1}{2}\|\mathbf{X}_2\mathbf{h}_1 - \mathbf{X}_1\mathbf{h}_2\|_2^2 + \lambda'(|\mathbf{h_1}| + |\mathbf{h_2}|)$$

$$\text{S.T.} \quad h_1(0) = 1$$

➢ Entire path of solutions (piecewise linear w.r.t λ') + cross validation

(M. R. Osborne *et al.*, 2000; B. Efron *et al.*, 2003; D. M. Malioutov *et al.*, 2005)

➢ Our proposal: learn the optimal λ' in a Bayesian framework

(*Evidence maximization*: D. J. C. MacKay, 1992;
Empirical Bayes: E. I. George *et al.*, 2000;
Sparse Bayesian learning (*RVM*): M. E. Tipping, 2001)

Probabilistic model:

$$P(\mathbf{X}_2\mathbf{h}_1 - \mathbf{X}_1\mathbf{h}_2|\mathbf{h}_1, \mathbf{h}_2, \sigma^2) \sim N(0, \sigma^2\mathbf{I})$$

$$P(\mathbf{h}_1, \mathbf{h}_2|\lambda) \propto \exp\{-\lambda(|\mathbf{h}_1| + |\mathbf{h}_2|)\}, \quad h_1(0) = 1$$

$$\lambda' = \sigma^2\lambda$$

Maximizing marginal likelihood

$$\lambda'^* = {\sigma^2}^* \lambda^*$$

$$\begin{aligned} {\sigma^2}^*, \lambda^* &= \arg\max_{\sigma^2, \lambda} [P(\mathbf{X}_2\mathbf{h}_1 - \mathbf{X}_1\mathbf{h}_2 | \sigma^2, \lambda)] \\ &= \int_{h_1(0)=1} P(\mathbf{X}_2\mathbf{h}_1 - \mathbf{X}_1\mathbf{h}_2 | \mathbf{h}_1, \mathbf{h}_2, \sigma^2) P(\mathbf{h}_1, \mathbf{h}_2 | \lambda) \mathbf{dh}_1 \mathbf{dh}_2 \end{aligned}$$

- Optimization: by EM updates

  --using a variational approach for computing expectations

  (Y. Lin *et al.*, 2006)

- The optimal regularization parameters: determined by filter statistics

  -- stationary in a given room

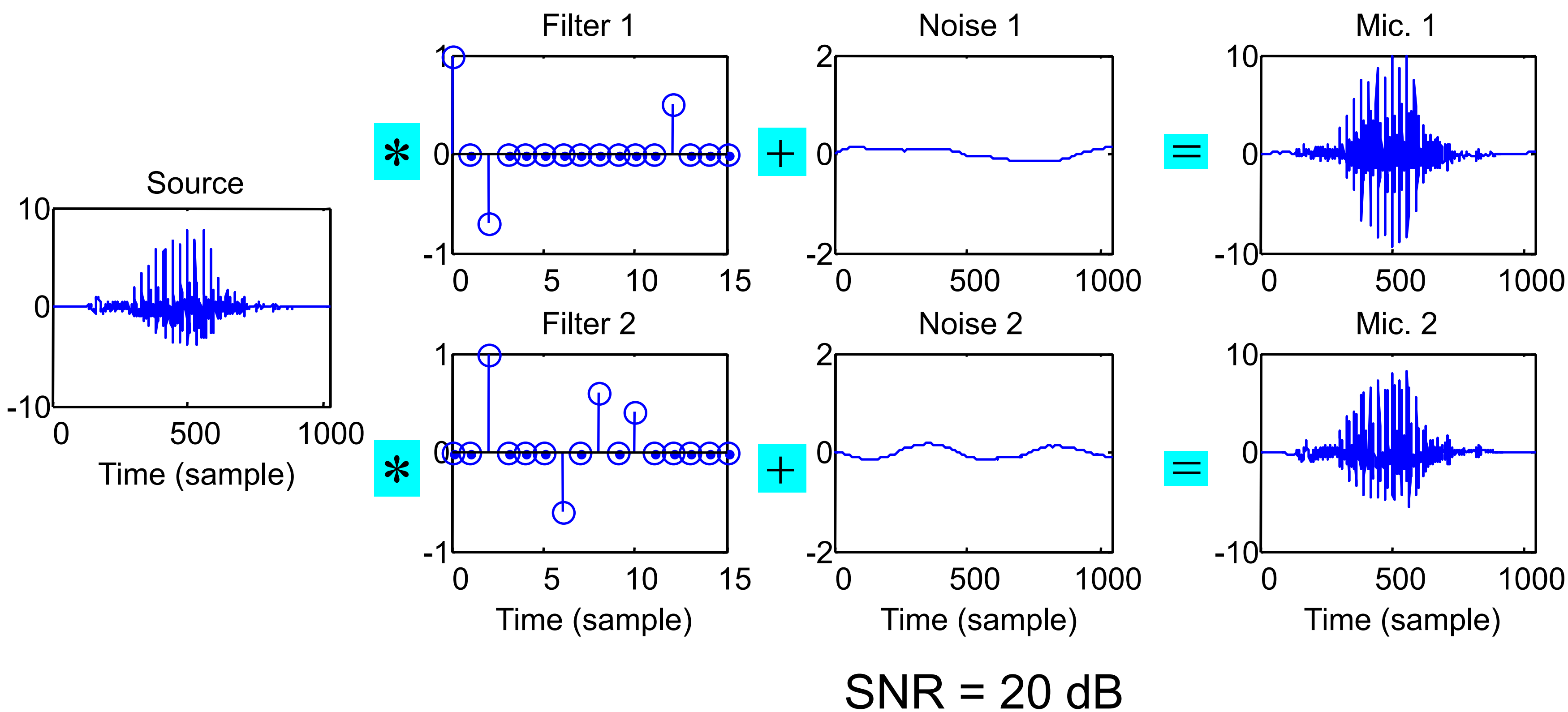
Source estimation

$$\mathbf{s}^* = \arg\min_{\mathbf{s}} \sum_{i=1}^{2} \|\mathbf{x}_i - \mathbf{s} * \mathbf{h}_i\|_2^2$$

- Yield ML-estimation if filter estimates are accurate;
- Solve efficiently in the frequency domain;
- Other methods: multiple-input/output inversion theorem (MINT).

(M. Miyoshi *et al.*, 1988)

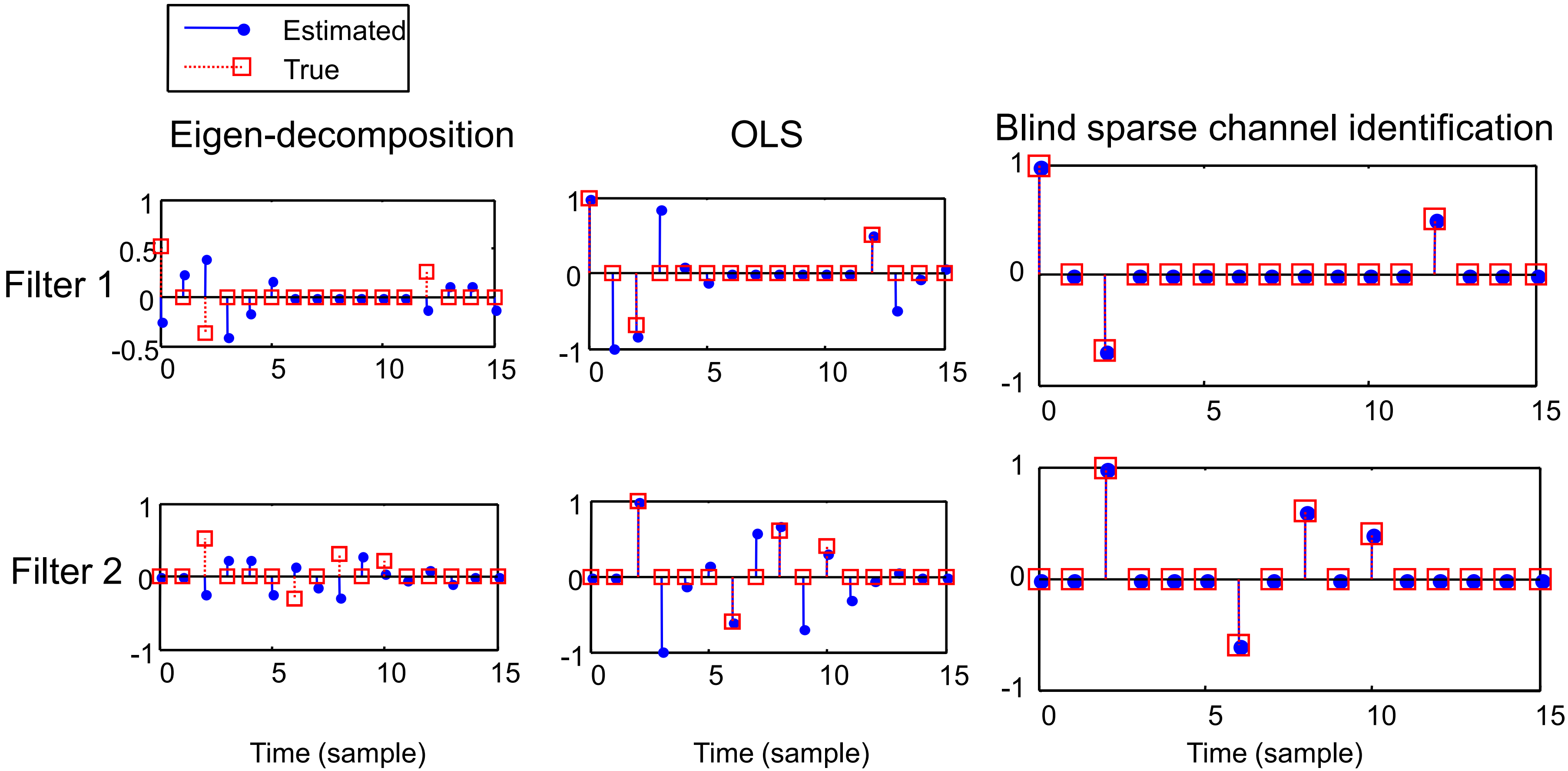
Simulations -- signals



- Blind channel identification:

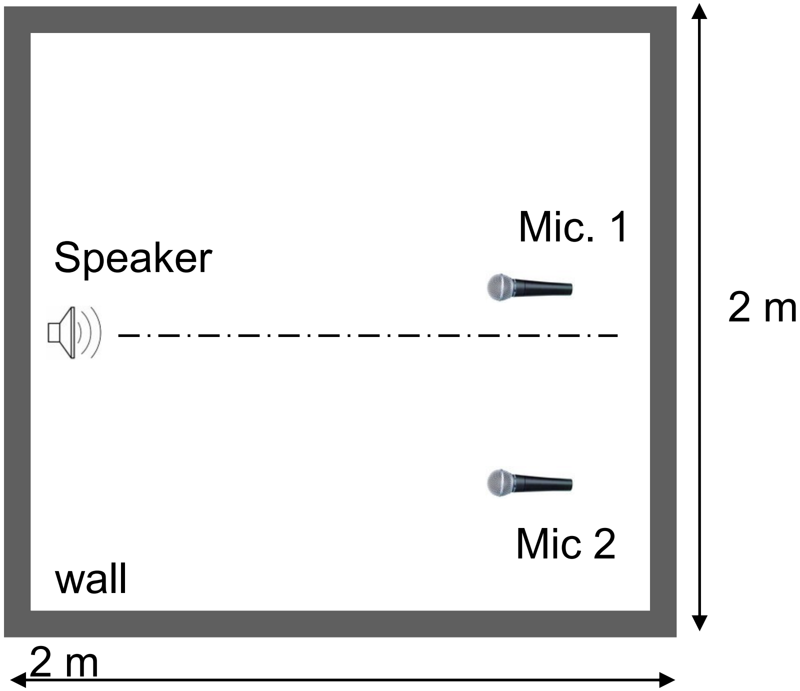
  -- estimates the filters given only microphone observations.

Simulations -- results

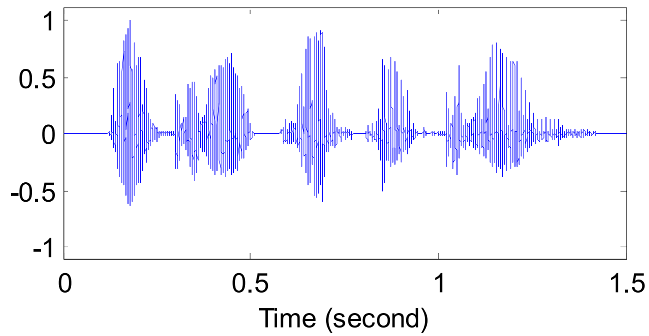


- OLS: yields better results than the eigenvalue decomposition approach.
- BSCI (LS + sparsity regularization): provides accurate filter estimates.

Experiments -- setup



Source:

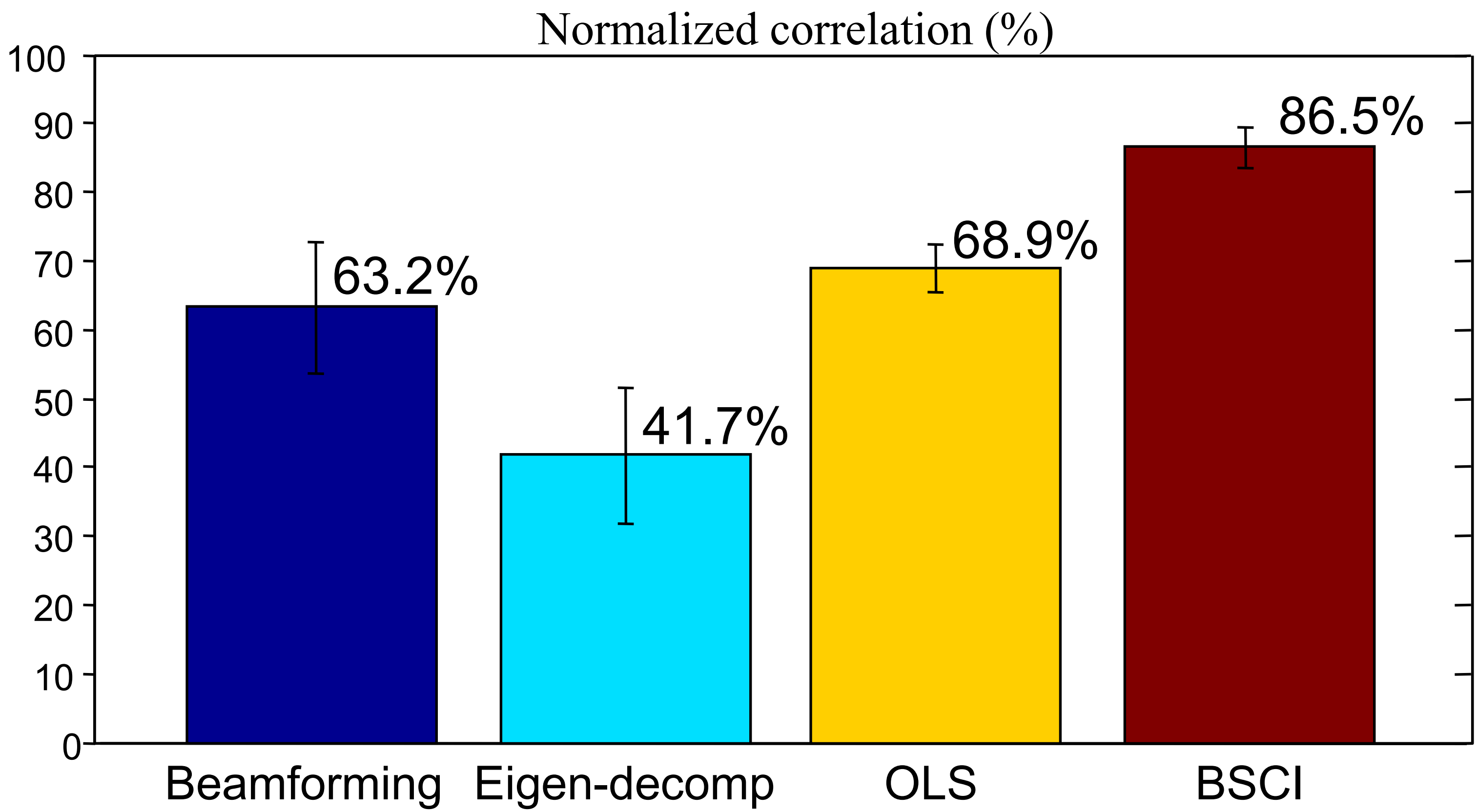


Data: 10 different speaker-microphone positions; **Sampling rate**: 16,000

Filter length: 1536 samples.

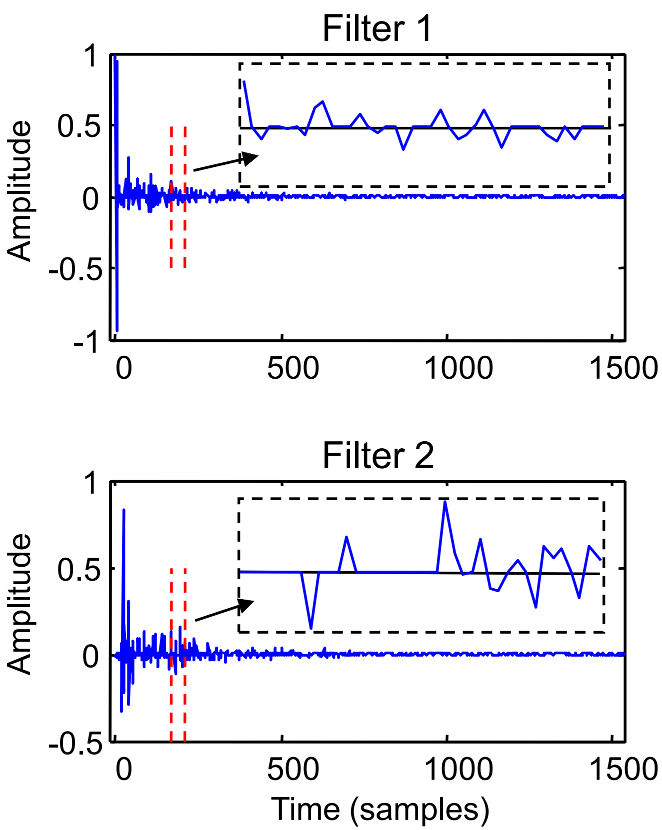
Anechoic chamber measurements: ground truth for the source.

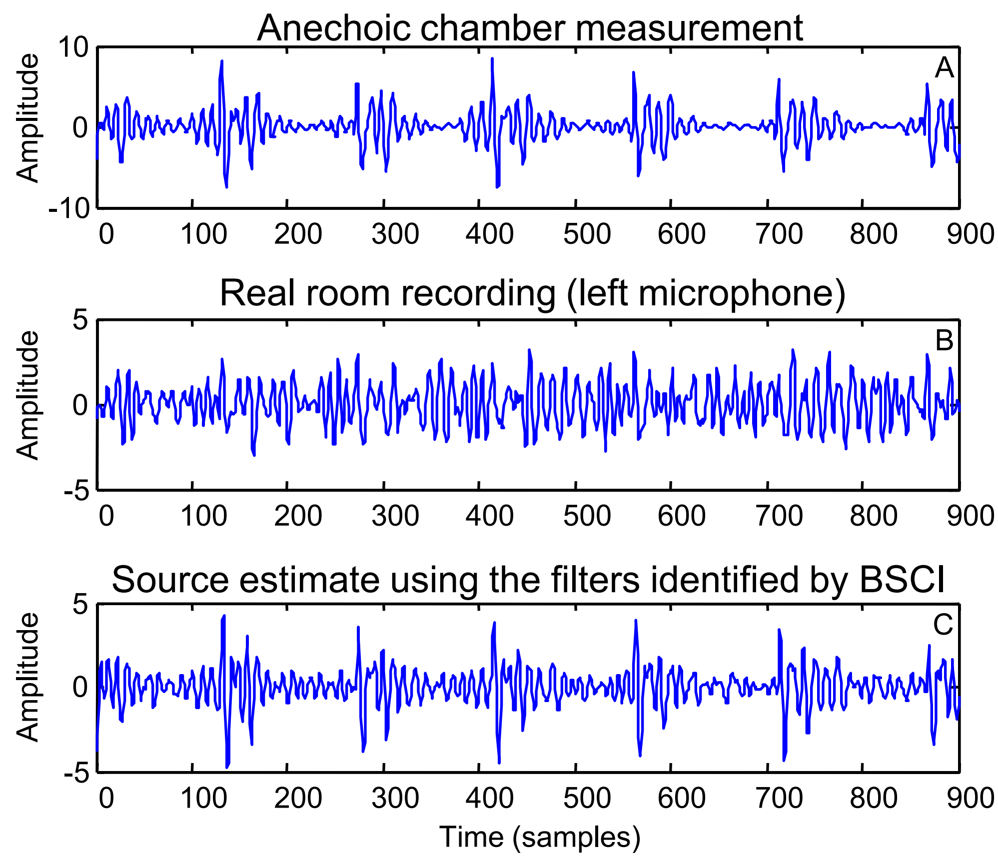
Results -- source estimates



- *OLS : yields better results than the eigenvalue decomposition approach.*
- *BSCI (LS + sparsity regularization):*
  -- provides source estimates with high fidelity.

Results -- filter and source estimates





About 2000 zero-elements

- Experimentally validate our sparse room impulse response model.

Summary

- **Blind sparse channel identification (BSCI) --- 3 key components:**

  1) Sparse room impulse response model;

  2) convex formulation;

  3) l_1-norm sparse Bayesian learning

  -- *for inferring the optimal sparsity regularization parameters*.

- **Our results demonstrate that the proposed BSCI has the potential to solve the speech dereverberation problem in real acoustic environments.**

Please come to our poster to listen to the sound examples…