2008 Tohoku-Section Joint Convention of Institutes of Electrical and Information Engineering, Japan

Blind Multichannel Identification in Acoustic Space using Numerous Channels

Takuma Okamoto, Yukio Iwaya and Yôiti Suzuki

Research Institute of Electrical Communication
Graduate School of Information Science
Tohoku University, Japan

Outline

1. Introduction

2. Blind Multichannel Identification

3. Computer Simulation

4. Conclusion

Introduction

Reconstruction techniques of Sound field information as cinemas and home theater

For high realistic sensation

The sound produced by these systems is, however, unrealistic!!

↓

Developing the system of recording, recognition and reproduction the sound field information in the real environment accurately

Sound field information
- Sound information
  - Sound source waveform data
- Field information
  - Reflected sound information
  - Sound source positions
  - Sound source number

Surrounding microphone array

Recording sound field information in the real environment

Surrounding microphone array @ RIEC

Spec of this room

- 5.18 × 3.38 × 2.22 (m)

- walls 20×2, 36×2
- ceiling 45

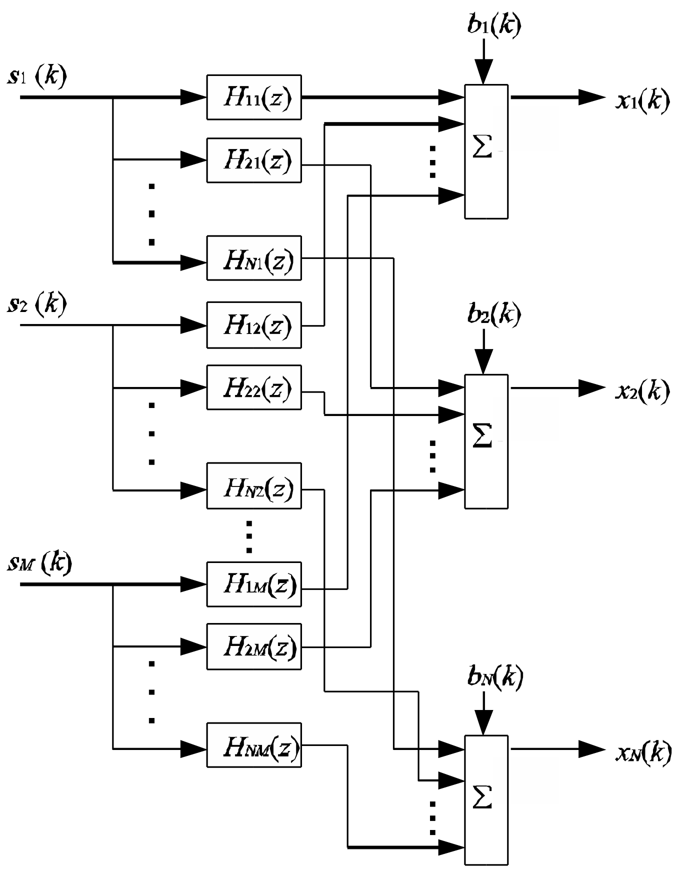
sum total 157 ch

- Installed 30 cm from all walls
- Separated from each other by 50 cm
- 48 kHz sampling (Synchronized 157 ch)

Estimating and extract sound field information using only the input signals of 157 microphones

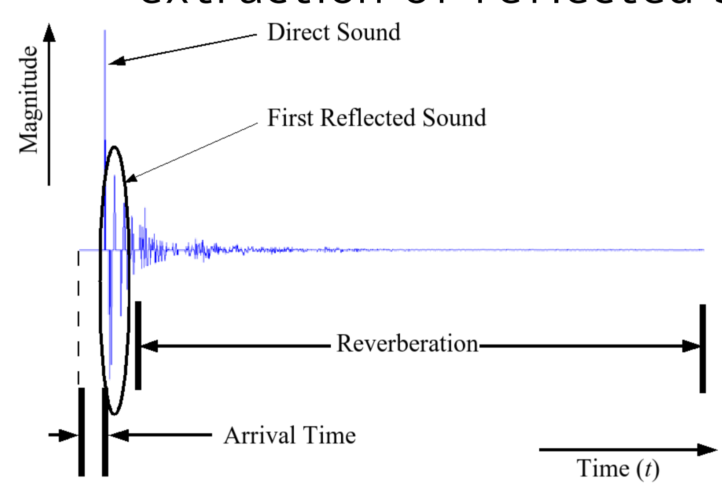
Accurately extraction of sound field information

Extraction of sound field information = Estimation of the room transfer functions (=impulse responses)



Multiple Input Multiple Output model (MIMO model)

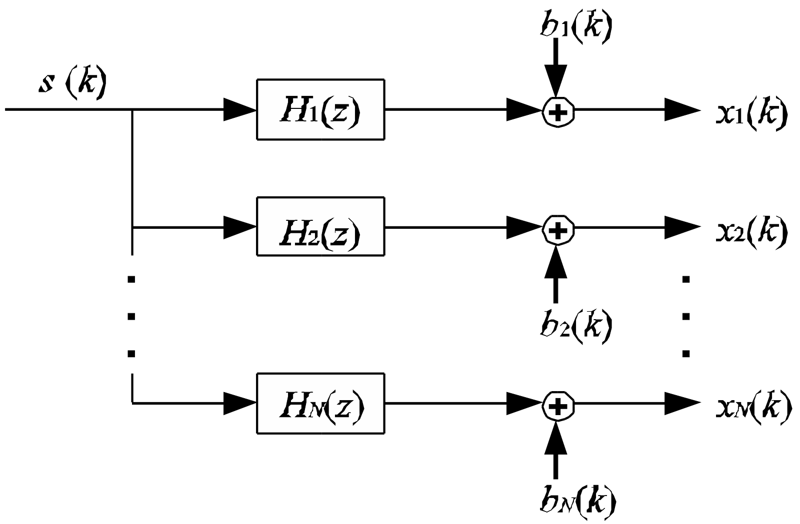
- Inverse filtering by MINT (M. Miyoshi *et al* 1989)
  → extraction of source waves
- Compute sound arrival time
  → extraction of source positions
- Source signal - input signals
  → extraction of reflected sound



Blind multichannel identification using 157 input signals

Estimation of the impulse responses from only input channels

Blind multichannel Identification (BMI) Y. Sato 1975



Single Input Multiple Output model
(SIMO model)

FNMLMS

(Frequency-domain normalized multichannel LMS)

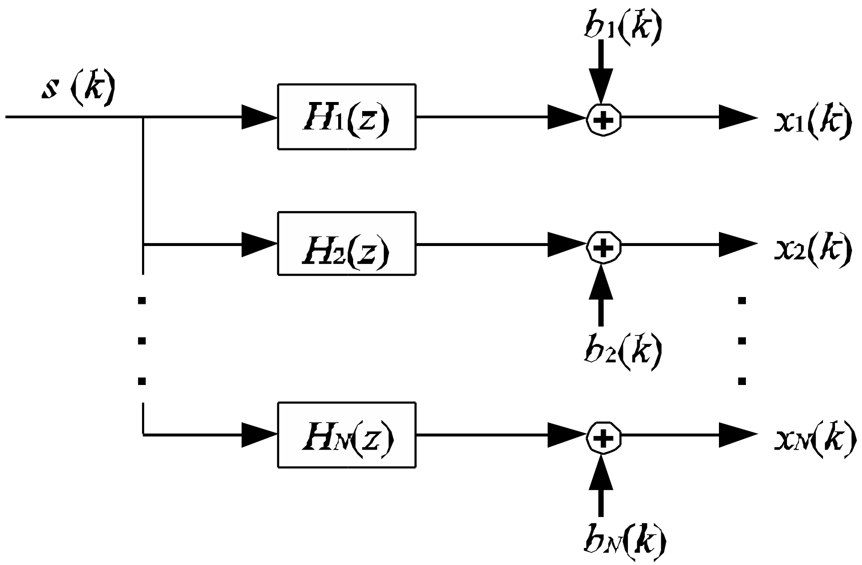
Y. Huang *et al* 2003

applied to BMI for estimating impulse responses in SIMO model

Y. Huang *et al* have used only three input signals (= much fewer than for our system)

Investigation of the performance of FNMLMS using 157 input signals

Blind multichannel identification based on second order statistics



Single Input Multiple Output model
(SIMO model)

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_N \end{bmatrix} \mathbf{s} + \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_M \end{bmatrix}$$

The output signal of i-th channel $x_i(n)$

$$x_i(n) = h_i * s(n) + b_i, \ \ i = 1, 2, \cdots, M \qquad (1)$$

In vector form

$$\mathbf{x}_i(n) = \mathbf{H}_i \cdot \mathbf{s}(n) + \mathbf{b}_i$$

$$\mathbf{x}_i(n) = [\ x_i(n) \quad x_i(n-1) \quad \cdots \quad x_i(n-L+1)\ ]^T$$

$$\mathbf{H}_i = \begin{bmatrix} h_{i,0} & h_{i,1} & \cdots & h_{i,L-1} & 0 & \cdots & 0 \\ 0 & h_{i,0} & \cdots & h_{i,L-2} & h_{i,L-1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_{i,0} & h_{i,1} & \cdots & h_{i,L-1} \end{bmatrix}$$

$$\mathbf{s}(n) = [\ s(n) \quad s(n-1) \quad \cdots \quad s(n-L+1) \quad \cdots \quad s(n-2L+2)\ ]^T$$

$$\mathbf{b}_i(n) = [\ b_i(n) \quad b_i(n-1) \quad \cdots \quad b_i(n-L+1)\ ]^T$$

$$\mathbf{h}_i = [\ h_{i,0} \quad h_{i,1} \quad \cdots \quad h_{i,L-1}\ ]^T$$

Time-domain multichannel LMS

From eq. (1)

$$x_i * h_j = s * h_i * h_j = x_j * h_i,\ i,j = 1,2,\cdots,M,\ i \neq j, \quad (2)$$

Therefore,

$$\mathbf{x}_i^T(n)\mathbf{h}_j = \mathbf{x}_j^T(n)\mathbf{h}_i,\ i,j = 1,2,\cdots,M,\ i \neq j,$$

and,

$$\mathbf{R}_{x_i x_i}\mathbf{h}_j = \mathbf{R}_{x_i x_j}\mathbf{h}_i,\ i,j = 1,2,\cdots,M,\ i \neq j,$$

※ $\mathbf{R}_{x_i x_j} = \mathrm{E}\{\mathbf{x}_i(n)\mathbf{x}_j^T(n)\}$

$$\sum_{i=1,i\neq j}^{M} \mathbf{R}_{x_i x_i}\mathbf{h}_j = \sum_{i=1,i\neq j}^{M} \mathbf{R}_{x_i x_j}\mathbf{h}_i,\ i,j = 1,2,\cdots,M$$

$$\mathbf{R}\mathbf{h} = \mathbf{0}$$

$$\mathbf{R} = \begin{bmatrix} \sum_{i\neq 1}\mathbf{R}_{x_i x_i} & -\mathbf{R}_{x_2 x_1} & \cdots & -\mathbf{R}_{x_M x_1} \\ -\mathbf{R}_{x_1 x_2} & \sum_{i\neq 2}\mathbf{R}_{x_i x_i} & \cdots & -\mathbf{R}_{x_M x_2} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{R}_{x_1 x_M} & -\mathbf{R}_{x_2 x_M} & \cdots & \sum_{i\neq M}\mathbf{R}_{x_i x_i} \end{bmatrix}$$

$$\mathbf{h} = [\ \mathbf{h}_1^T \quad \mathbf{h}_2^T \quad \cdots \quad \mathbf{h}_M^T\ ]^T$$

Time-domain multichannel LMS (2)

Estimated impulse responses

$$\hat{\mathbf{h}} = [\ \hat{\mathbf{h}}_1^T \quad \hat{\mathbf{h}}_2^T \quad \cdots \quad \hat{\mathbf{h}}_M^T\ ]^T$$

An error signal

$$\epsilon_{ij}(n) = \begin{cases} \mathbf{x}_i^T(n)\hat{\mathbf{h}}_j - \mathbf{x}_j^T(n)\hat{\mathbf{h}}_i, & i \neq j, i, j, = 1, 2, \cdots, M \\ 0, & i = j, i, j, = 1, 2, \cdots, M \end{cases} \qquad (3)$$

The cost function

$$J(n) = \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \epsilon_{ij}^2(n) \qquad (4)$$

$$\tilde{\mathbf{R}} = \begin{bmatrix} \sum_{i\neq 1} \tilde{\mathbf{R}}_{x_i x_i} & -\tilde{\mathbf{R}}_{x_2 x_1} & \cdots & -\tilde{\mathbf{R}}_{x_M x_1} \\ -\tilde{\mathbf{R}}_{x_1 x_2} & \sum_{i\neq 2} \tilde{\mathbf{R}}_{x_i x_i} & \cdots & -\tilde{\mathbf{R}}_{x_M x_2} \\ \vdots & \vdots & \ddots & \vdots \\ -\tilde{\mathbf{R}}_{x_1 x_M} & -\tilde{\mathbf{R}}_{x_2 x_M} & \cdots & \sum_{i\neq M} \tilde{\mathbf{R}}_{x_i x_i} \end{bmatrix}$$

$$\mathbf{R}_{x_i x_j} = \mathrm{E}\{\mathbf{x}_i(n)\mathbf{x}_j^T(n)\},\ i, j = 1, 2, \cdots, M$$

Constraint

$$\hat{\mathbf{h}} = \arg\min \mathrm{E}\{J(n)\},\ \text{subject to} \|\hat{\mathbf{h}}\| = 1$$

Multichannel LMS

$$\hat{\mathbf{h}}(n+1) = \frac{\hat{\mathbf{h}}(n) - 2\mu[\tilde{\mathbf{R}}(n)\hat{\mathbf{h}}(n) - J(n)\hat{\mathbf{h}}]}{\|\hat{\mathbf{h}}(n) - 2\mu[\tilde{\mathbf{R}}(n)\hat{\mathbf{h}}(n) - J(n)\hat{\mathbf{h}}]\|}$$

Frequency-domain multichannel LMS

By taking advantage of the computational efficiency of the FFT
→ a convolution of two signals can be quickly calculated

Discrete Fourier transform processes a time sequence like a filter bank, which orthogonalizes the data,

↓

the coefficients of a frequency-domain adaptive filter can converge independently or even uniformly if the update is normalized properly

Frequency-domain multichannel LMS (FMLMS)

Calculating a convolution by FFT and the cost function is frequency-domain

Frequency-domain normalized multichannel LMS (FNMLMS)

Introducing the forgetting factor to FMLMS and accelerate the convergence

Computer simulation

Simulation conditions

sound sourse $s(n)$: Gaussian white noise

impulse response h_i : 64 taps of FIR filter (made by image method)

numbers of channels M : 3 and 157

step size parameter ρ_f : 1.0

Image method J. B. Allen *et al* 1979

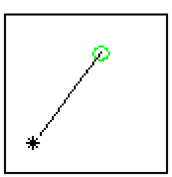


Figure 1

Direct sound

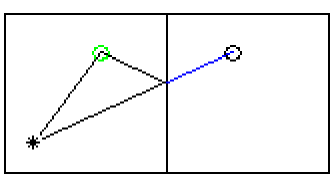


Figure 2

First reflected sound

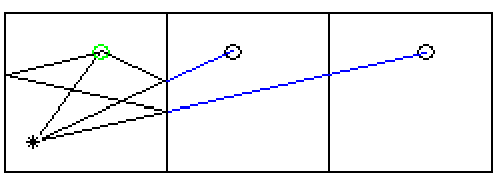


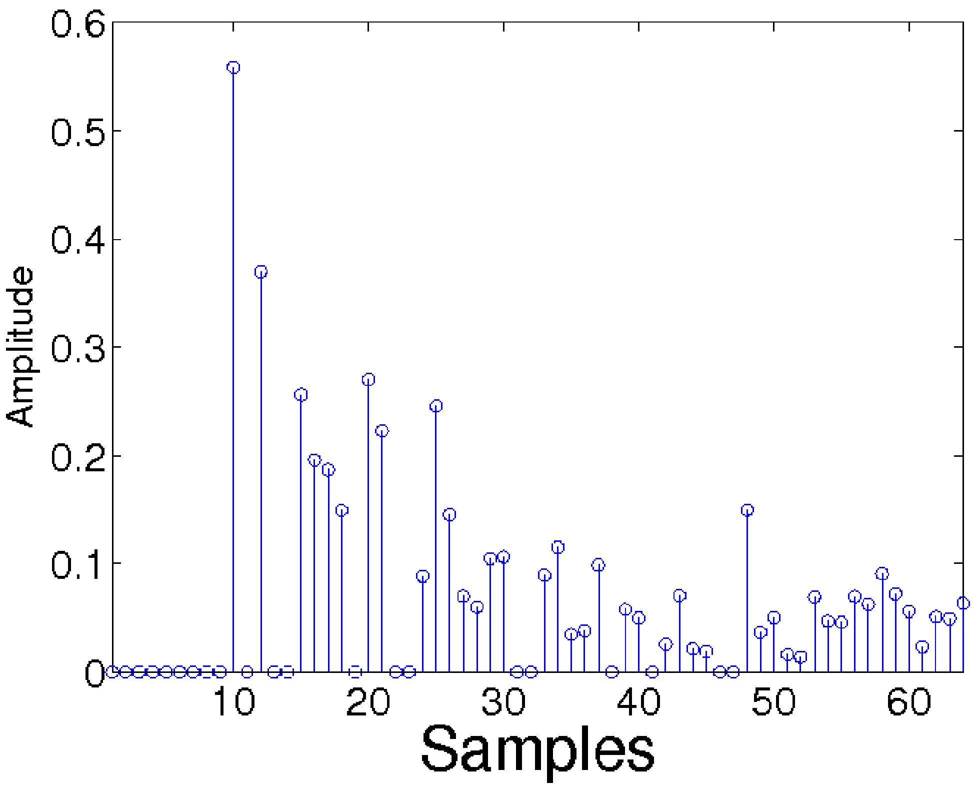
Figure 3

Second reflected sound

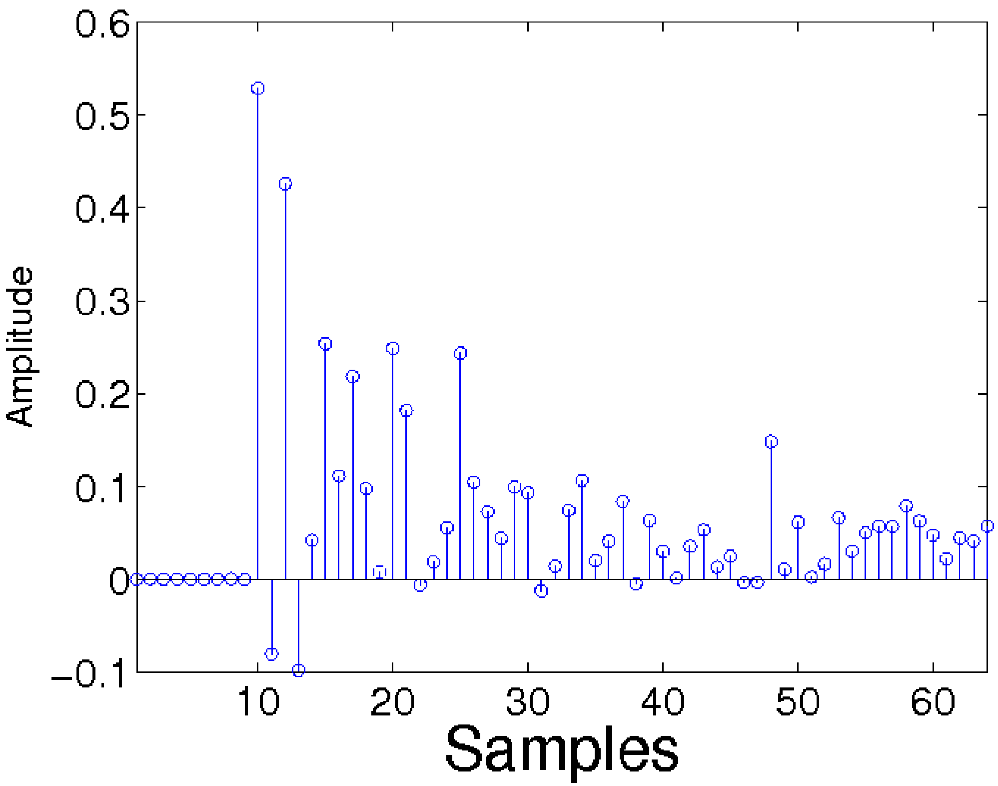
Figure 4

Real and image source positions

Simulation results

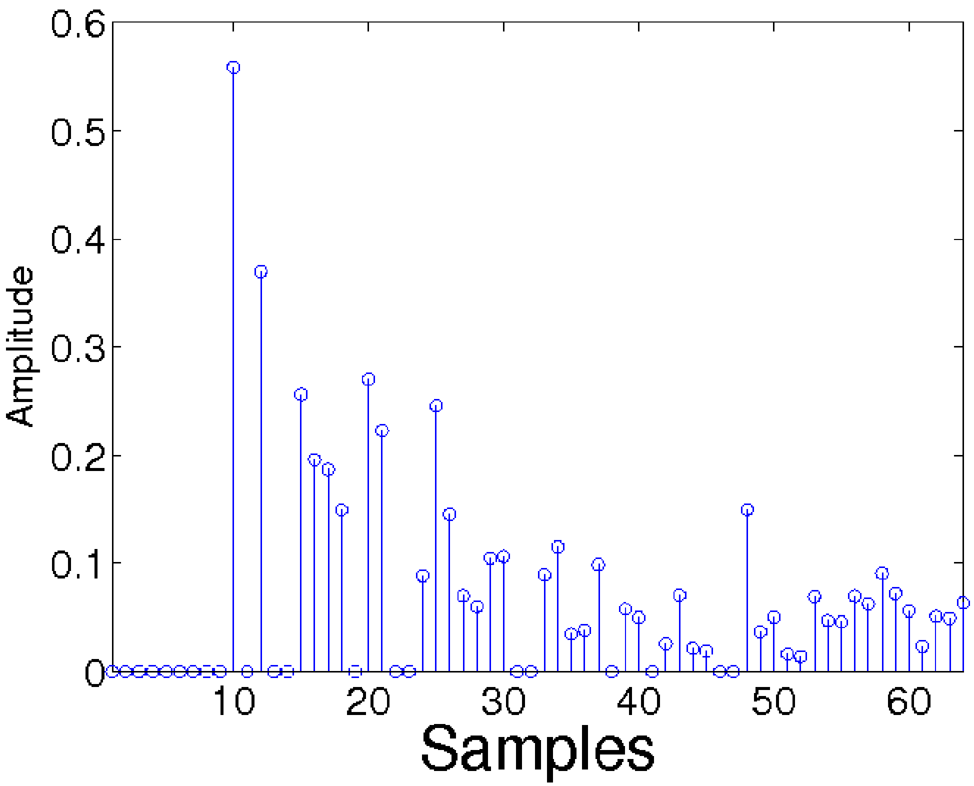


Generated Impulse Response (channel 1)

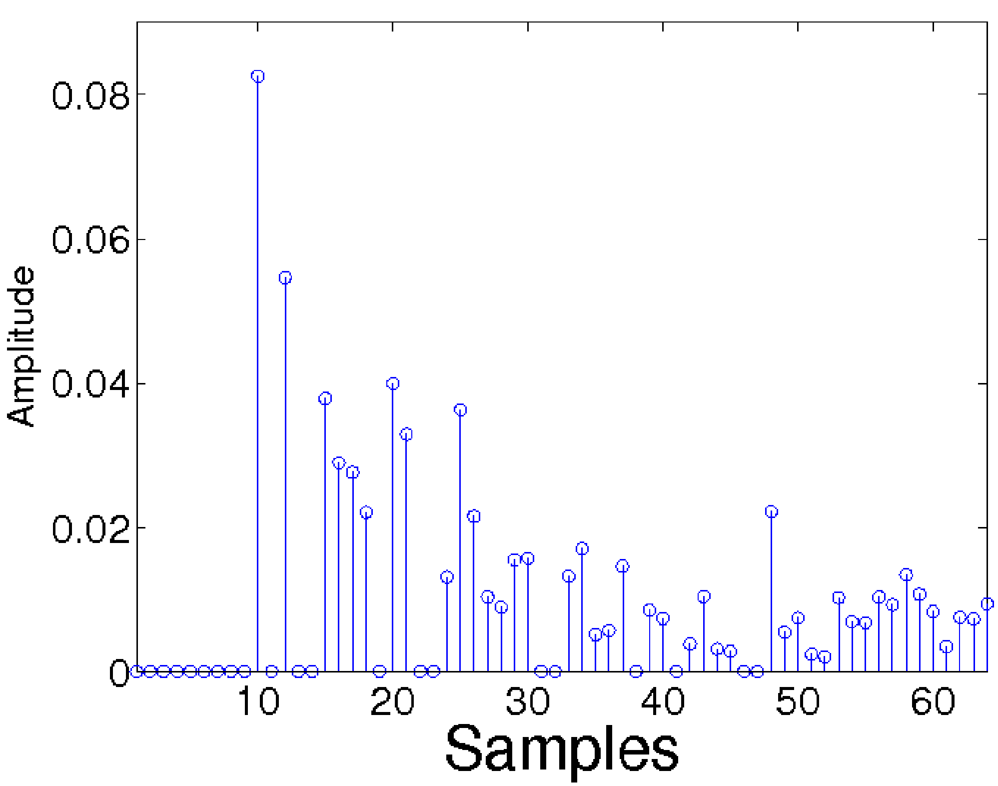


Estimated Impulse Response (N = 3)

Simulation results (2)



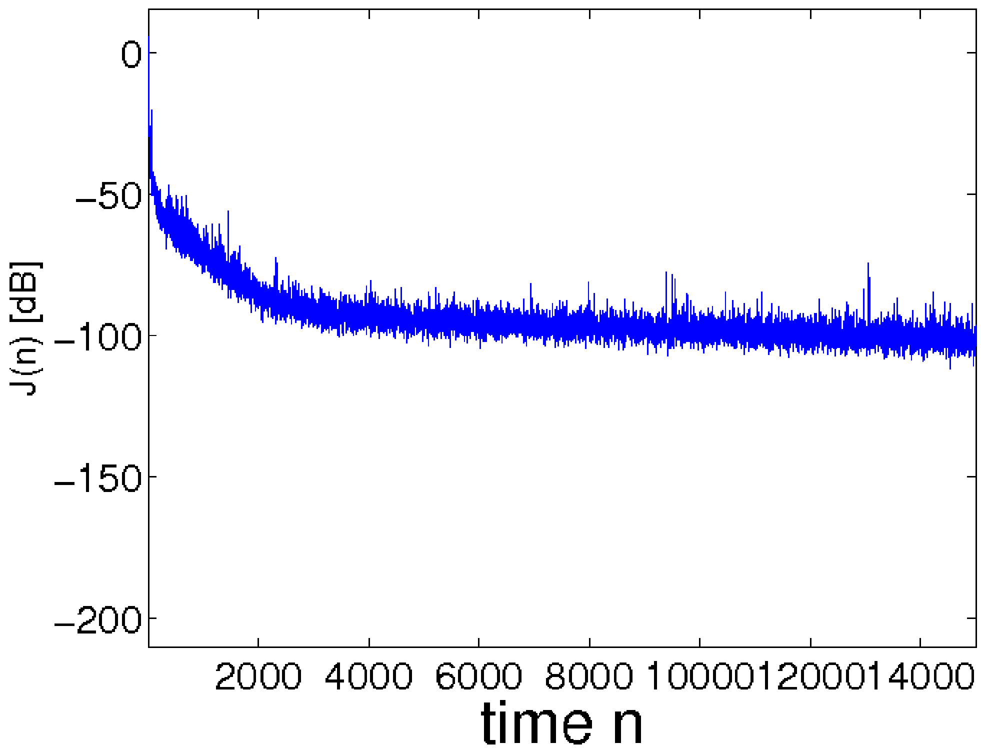
Generated Impulse Response (channel 1)



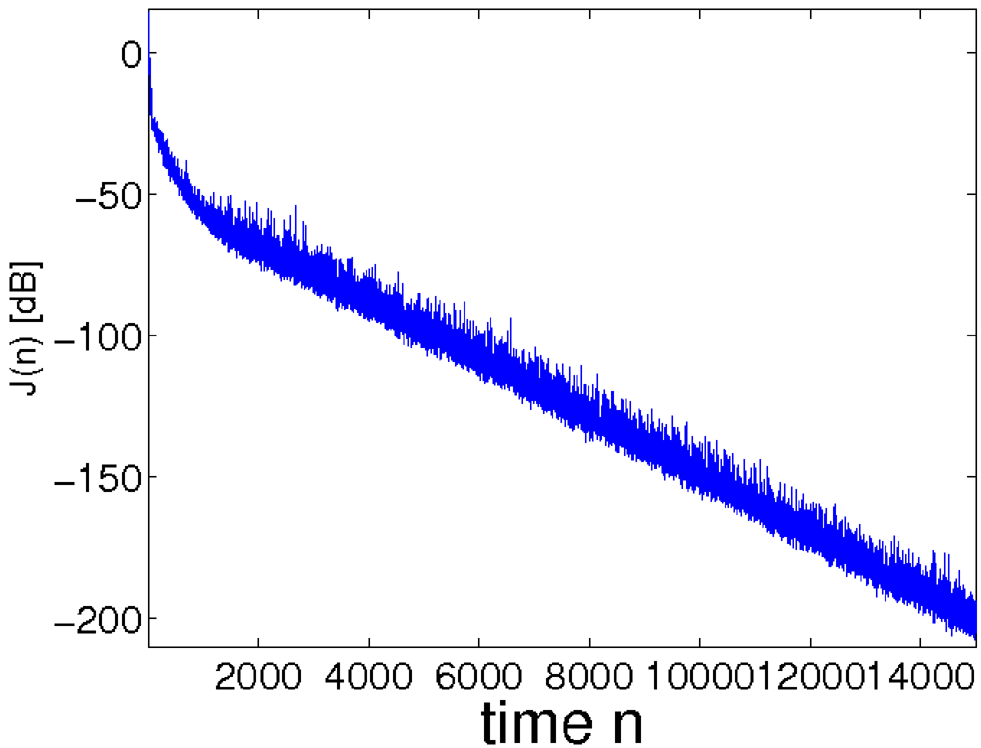
Estimated Impulse Response (N = 3)

Simulation results (3)

The cost function $$J(n) = \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \epsilon_{ij}^2(n) \qquad (4)$$



Generated Impulse Response (channel 1)



Estimated Impulse Response (N = 3)

Conclusion

Comparing the performance of FNMLMS using 3 and 157 channels

Using 3 channels

Local solution is computed for few channels

Using 157 channels

True solution is computed

FNMLMS using 157 channels is superior to that using 3 channels !!