Christian-Albrechts-Universität zu Kiel

Adaptive Filters – Adaptation Control

Gerhard Schmidt

Christian-Albrechts-Universität zu Kiel
Faculty of Engineering
Electrical Engineering and Information Technology
Digital Signal Processing and System Theory

Contents of the Lecture

Today:

Adaptation Control:

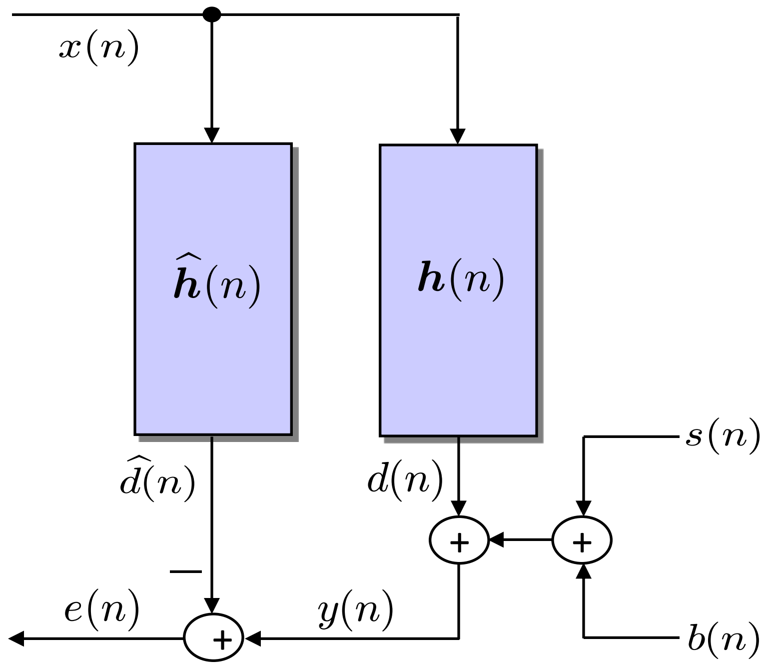
- Introduction and Motivation
- Prediction of the System Distance
- Optimum Control Parameters
- Estimation Schemes
- Examples

Basics

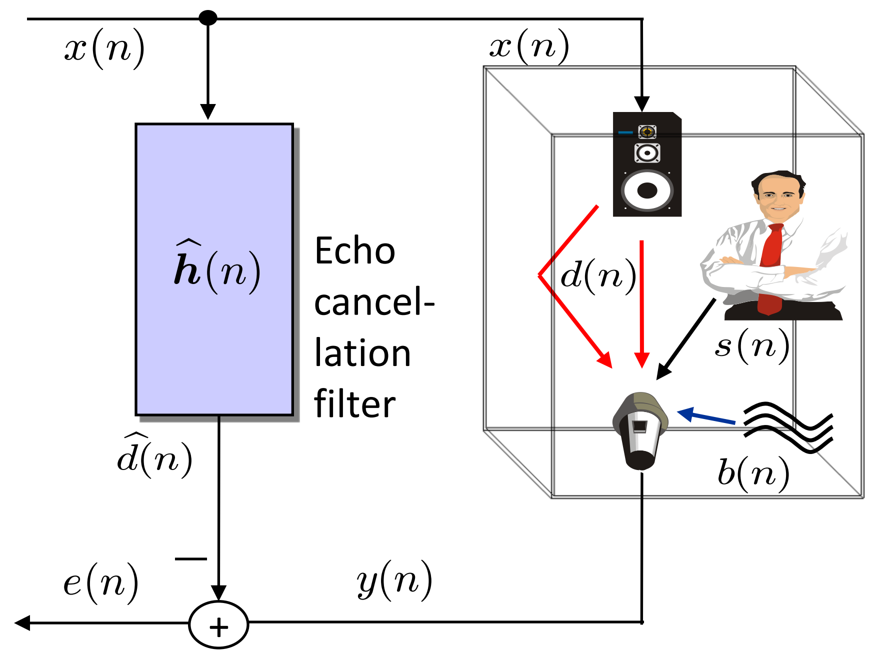
Objective:

Remove those components in the microphone signal that originate from the remote communication partner!

Model:



Application example:



Basic Approach

Model:

The loudspeaker-enclosure-microphone (LEM) system is modelled as a linear (only slowly changing) system with finite memory.

Approach:

Cancelling acoustic echoes by means of an adaptive filter with $N = 1024$ coefficients, operating at a sample rate $f_\mathrm{s} = 8$ kHz. For the adaptation of the filter the NLMS algorithm should be used.

Advantages and disadvantages:

+ In contrast to former approaches (loss controls) simultaneous speech activity in both communication directions is possible now.
+ The NLMS algorithm is a robust and computationally efficient approach.
– Compared to former solutions more memory and a larger computational load are required.
– Stability can not be guaranteed.

NLMS-Algorithm

Computation of the error signal (output signal of the echo cancellation filter):

$$e(n) \;=\; y(n) - \Big[x(n),\, x(n-1),\, \ldots,\, x(n-N+1)\Big] \begin{bmatrix} \widehat{h}_0(n) \\ \widehat{h}_1(n) \\ \vdots \\ \widehat{h}_{N-1}(n) \end{bmatrix}$$

Recursive computation of the norm of the excitation signal vector

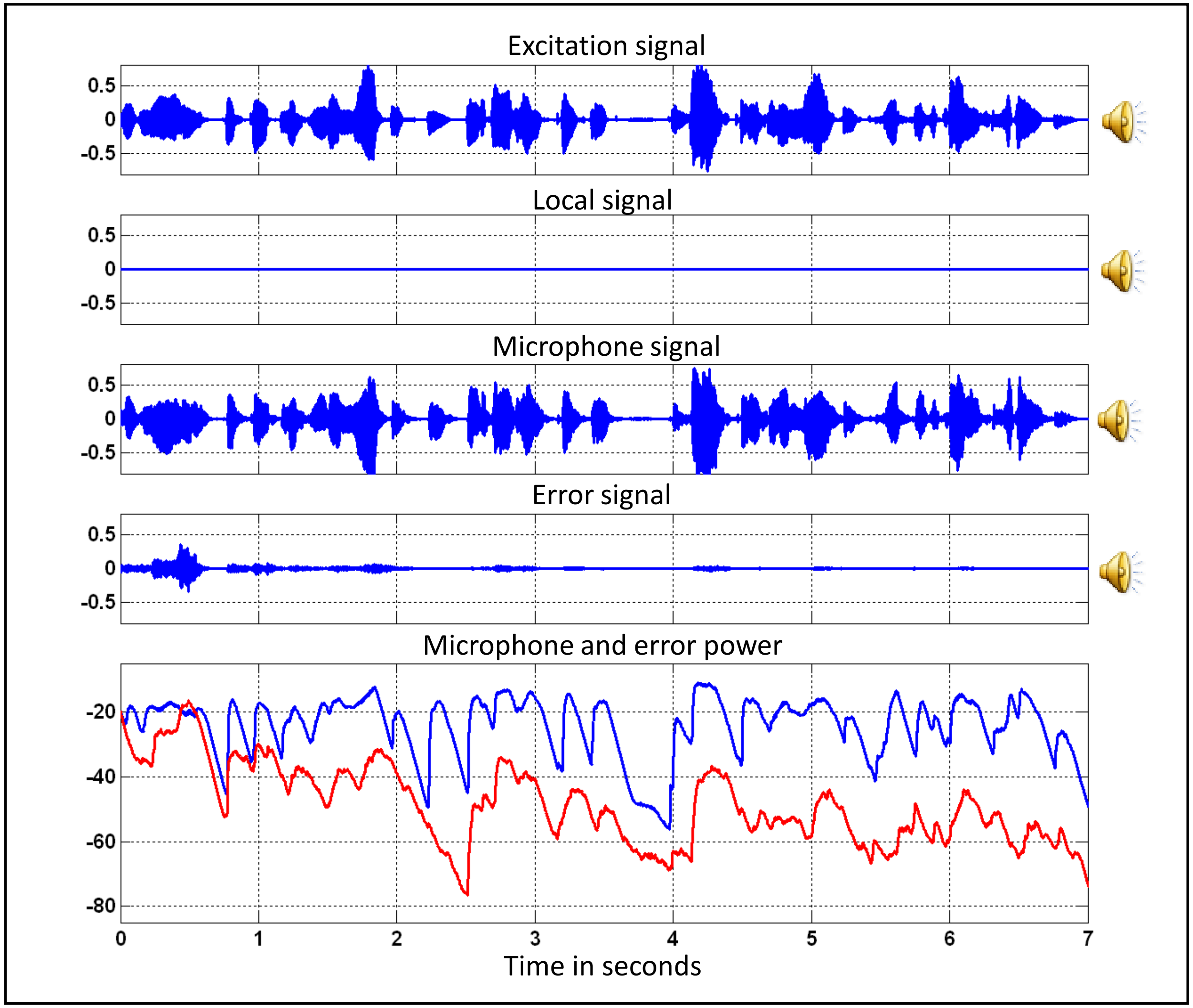
$$\big\|\boldsymbol{x}(n)\big\|^2 \;=\; \big\|\boldsymbol{x}(n-1)\big\|^2 - x^2(n-N) + x^2(n)$$

Adaptation of the filter vector:

$$\begin{bmatrix} \widehat{h}_0(n+1) \\ \widehat{h}_1(n+1) \\ \vdots \\ \widehat{h}_{N-1}(n+1) \end{bmatrix} = \begin{bmatrix} \widehat{h}_0(n) \\ \widehat{h}_1(n) \\ \vdots \\ \widehat{h}_{N-1}(n) \end{bmatrix} + \frac{\mu\, e(n)}{\big\|\boldsymbol{x}(n)\big\|^2} \begin{bmatrix} x(n) \\ x(n-1) \\ \vdots \\ x(n-N+1) \end{bmatrix}$$

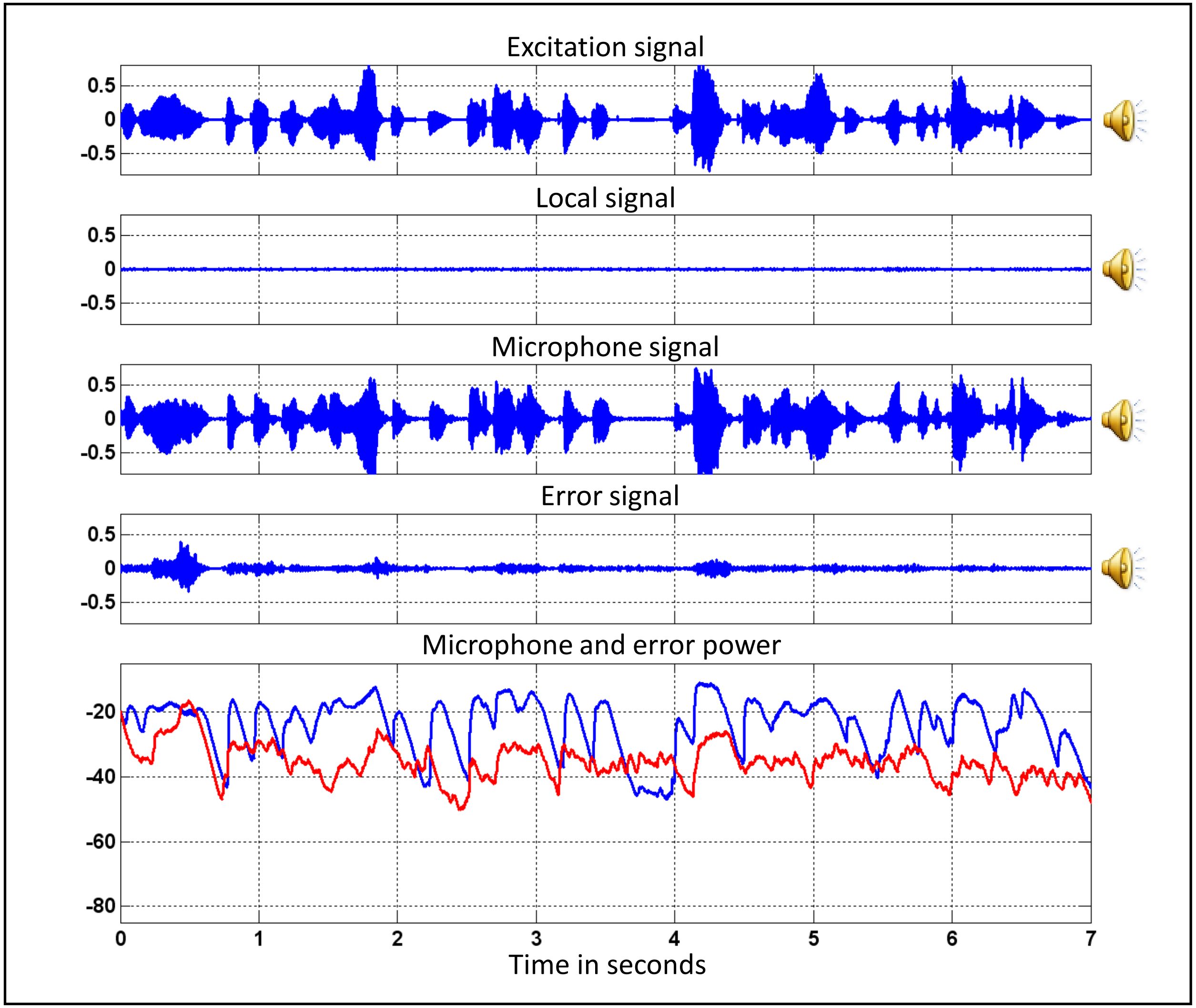
Convergence Examples – Part 1

Convergence without background noise and without local speech signals



Convergence Examples – Part 2

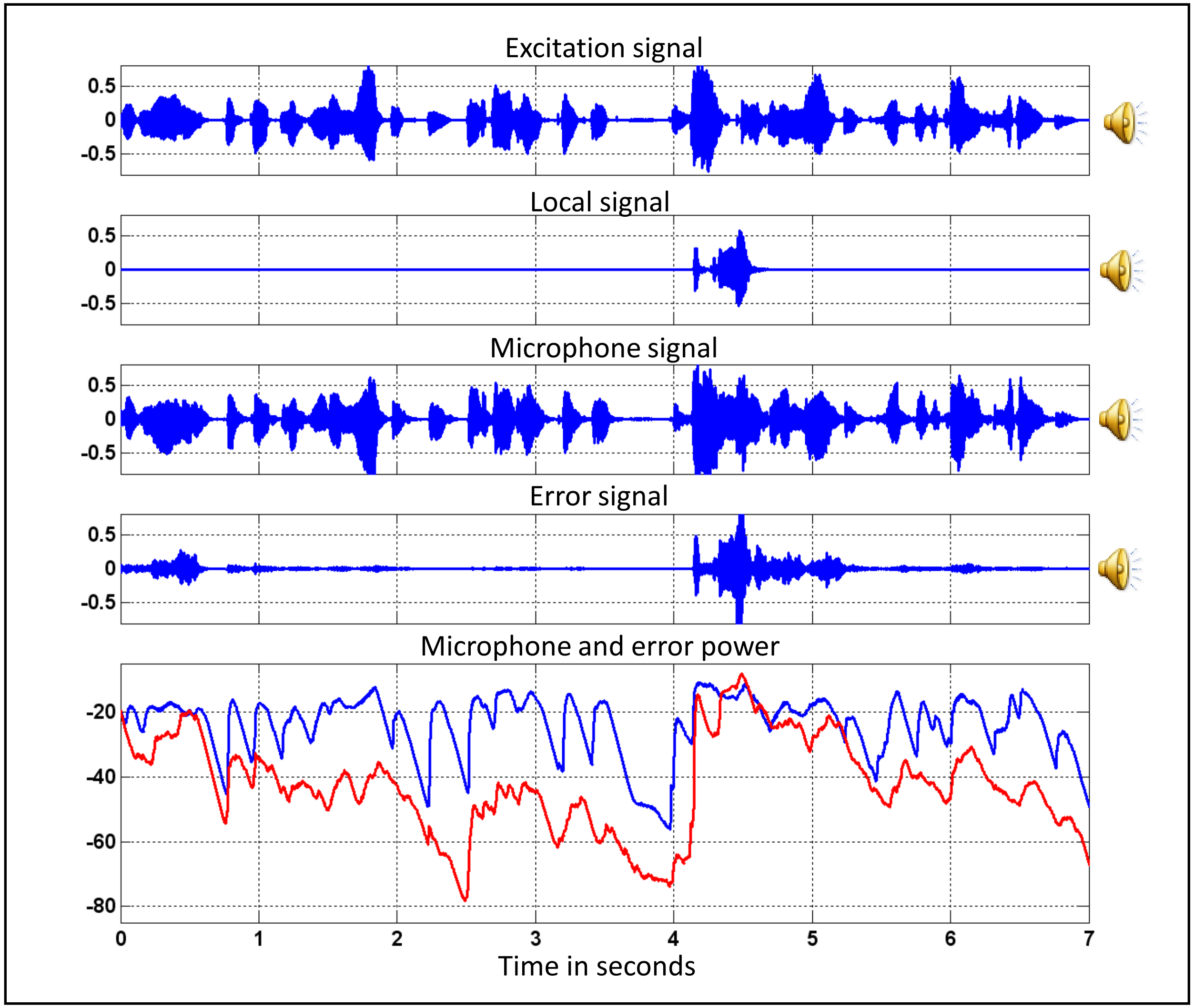
Convergence with background noise but without local speech signals



Convergence Examples – Part 3

Convergence without background noise but with local speech signals

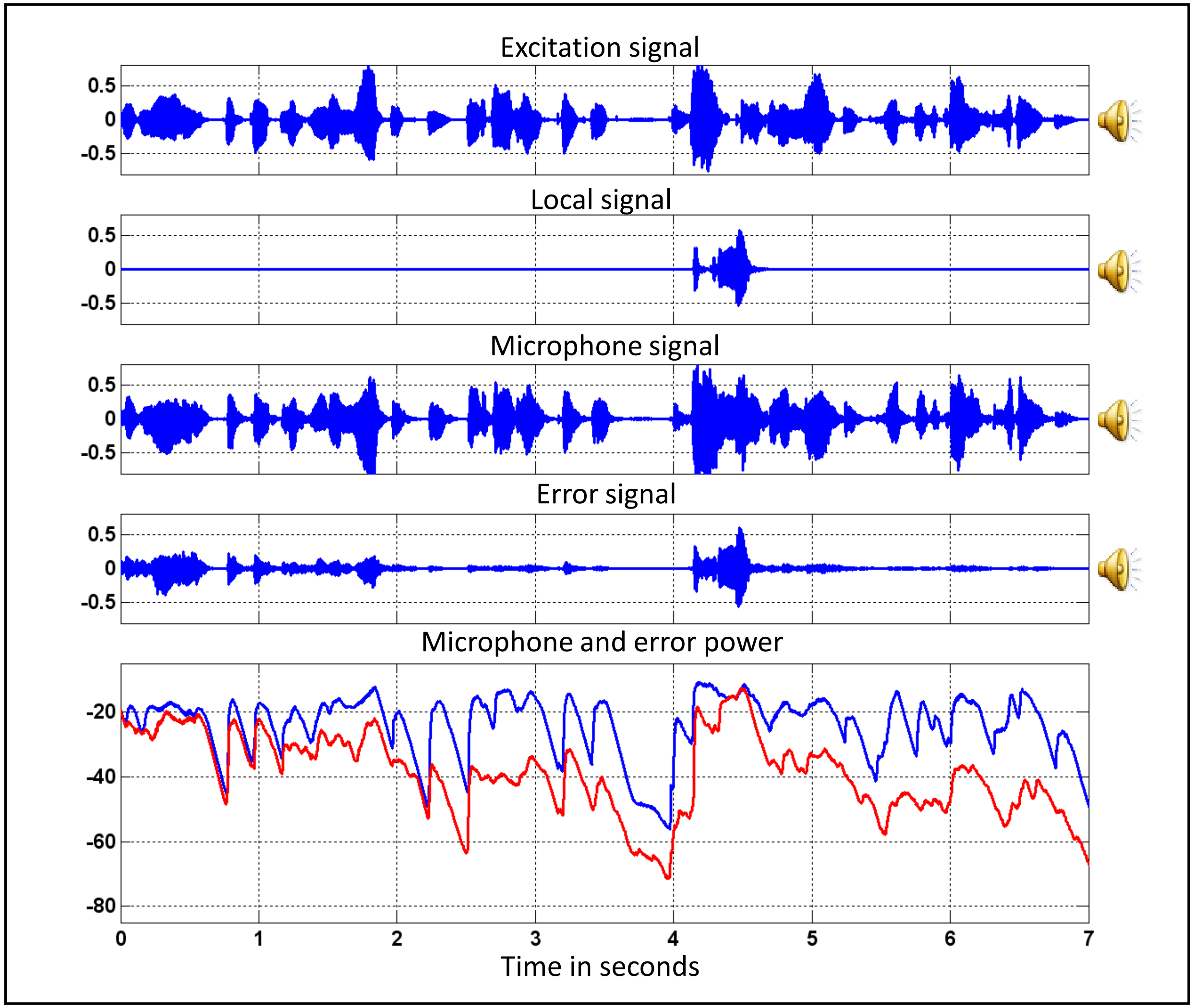
(step size = 1)



Convergence Examples – Part 4

Convergence without background noise but with local speech signals

(step size = 0.1)



Literature

Basic texts:

- E. Hänsler / G. Schmidt: *Acoustic Echo and Noise Control* – Chapter 7 (Algorithms for Adaptive Filters), Wiley, 2004
- E. Hänsler / G. Schmidt: *Acoustic Echo and Noise Control* – Chapter 13 (Control of Echo Cancellation Systems), Wiley, 2004

Further details:

- S. Haykin: *Adaptive Filter Theory* – Chapter 6 (Normalized Least-Mean-Square Adaptive Filters), Prentice Hall, 2002
- C. Breining, A. Mader: *Intelligent Control Strategies for Hands-Free Telephones*, in E. Hänsler, G. Schmidt, Topics on Acoustic Echo and Noise Control – Chapter 8, Springer, 2006

Control Approaches – Part 1

Scalar control approach:

$$\widehat{\boldsymbol{h}}(n+1) \;=\; \widehat{\boldsymbol{h}}(n) + \underbrace{\mu(n)}_{\text{Step size}} \frac{\boldsymbol{x}(n)\, e(n)}{\left\|\boldsymbol{x}(n)\right\|^2 + \underbrace{\Delta(n)}_{\text{Regularization}}}$$

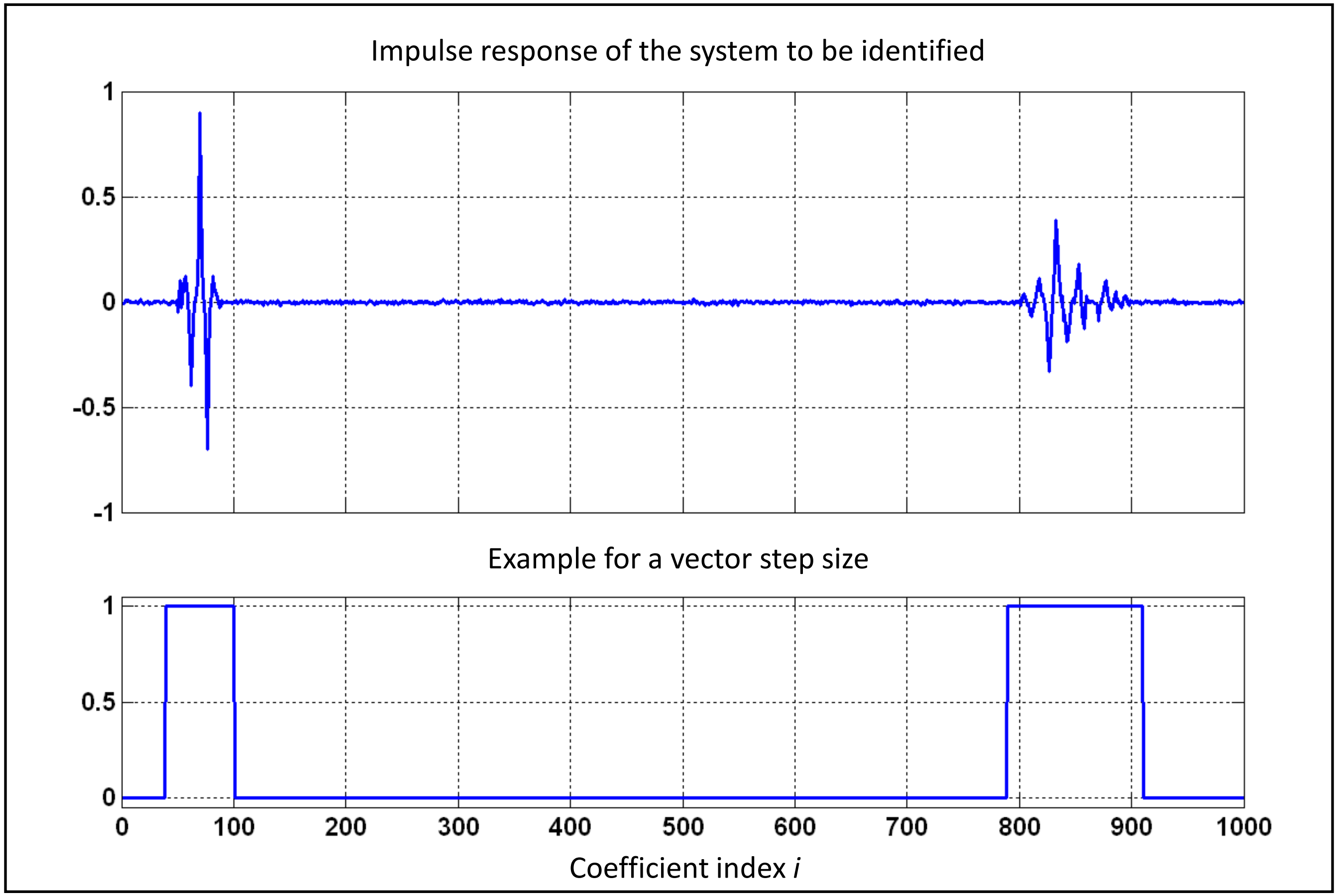
Vector control approach:

$$\widehat{\boldsymbol{h}}(n+1) \;=\; \widehat{\boldsymbol{h}}(n) + \begin{bmatrix} \mu_0(n) & 0 & \dots & 0 \\ 0 & \mu_1(n) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mu_{N-1}(n) \end{bmatrix} \frac{\boldsymbol{x}(n)\, e(n)}{\left\|\boldsymbol{x}(n)\right\|^2 + \Delta(n)}$$

Control Approaches – Part 2

Example for a sparse impulse response

For such systems a vector based control scheme can be advantageous.



Christian-Albrechts-Universität zu Kiel

How do we go on …

Problem
(echo cancellation performance during „double talk“)

Analysis of the average system distance
(taking local signals into account)

Derivation of an optimal step size
(using non-measurable signals)

Estimation of the non-measurable signal components
(leads to an implementable control scheme)

Solution
(robust echo cancellation due to step-size control)

Average System Distance – Part 1

Assumptions:

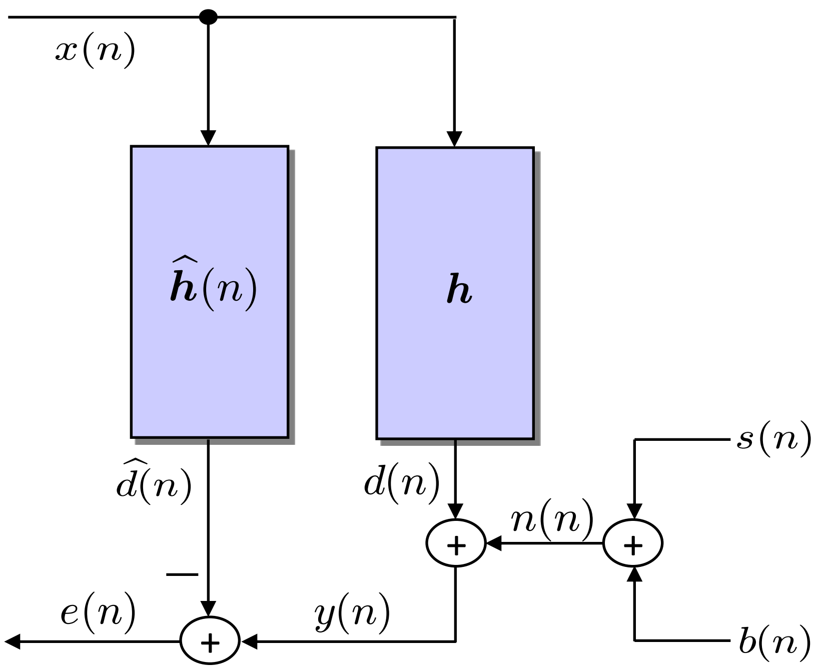
- Adaptation using the NLMS algorithm (only step-size controlled) :

$$\widehat{\boldsymbol{h}}(n+1) = \widehat{\boldsymbol{h}}(n) + \mu\,\frac{e(n)\,\boldsymbol{x}(n)}{\big\|\boldsymbol{x}(n)\big\|^2}$$

- White noise as excitation and (stationary) distortion:

$$r_{xx}(l) = \begin{cases} \sigma_x^2, & \text{if } l = 0,\\ 0, & \text{else.} \end{cases}$$

- Statistical independence between filter vector and excitation vector.
- Time-invariant system: $\boldsymbol{h}$



Definition of the average system distance:

$$\mathrm{E}\Big\{\big\|\boldsymbol{h}_\Delta(n)\big\|^2\Big\} = \mathrm{E}\Big\{\big\|\boldsymbol{h} - \widehat{\boldsymbol{h}}(n)\big\|^2\Big\}$$

Average System Distance – Part 2

… Derivation during the lecture …

Average System Distance – Part 3

Generic approach (control scheme with step size and regularization):

$$\widehat{\boldsymbol{h}}(n+1) = \widehat{\boldsymbol{h}}(n) + \mu \, \frac{\boldsymbol{x}(n)\, e(n)}{\|\boldsymbol{x}(n)\|^2 + \Delta}$$

Result:

$$\mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n+1)\|^2\right\} \approx \underbrace{\left(1 + \frac{\mu^2\, N\, \sigma_x^4}{(N\,\sigma_x^2+\Delta)^2} - \frac{2\,\mu\,\sigma_x^2}{N\,\sigma_x^2+\Delta}\right)}_{A(\mu,\Delta,\sigma_x^2,N)} \mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n)\|^2\right\} + \underbrace{\frac{\mu^2\, N\, \sigma_x^2}{(N\,\sigma_x^2+\Delta)^2}}_{B(\mu,\Delta,\sigma_x^2,N)} \sigma_n^2$$

$$= A(\mu,\Delta,\sigma_x^2,N)\, \mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n)\|^2\right\} + B(\mu,\Delta,\sigma_x^2,N)\,\sigma_n^2$$

Contraction parameter ***Expansion parameter***

Contraction and Expansion Parameters

$$\boxed{\mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n+1)\|^2\right\} \approx A(\mu,\Delta,\sigma_x^2,N)\,\mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n)\|^2\right\} + B(\mu,\Delta,\sigma_x^2,N)\,\sigma_n^2}$$

Contraction parameter $A(\mu,\Delta,\sigma_x^2,N)$ ***:***

- Range: $1-\dfrac{1}{N} \leq A(\mu,\Delta,\sigma_x^2,N) \leq \infty$
- Desired: as small as possible
- Determines the speed of convergence without distortions

Expansion parameter $B(\mu,\Delta,\sigma_x^2,N)$ ***:***

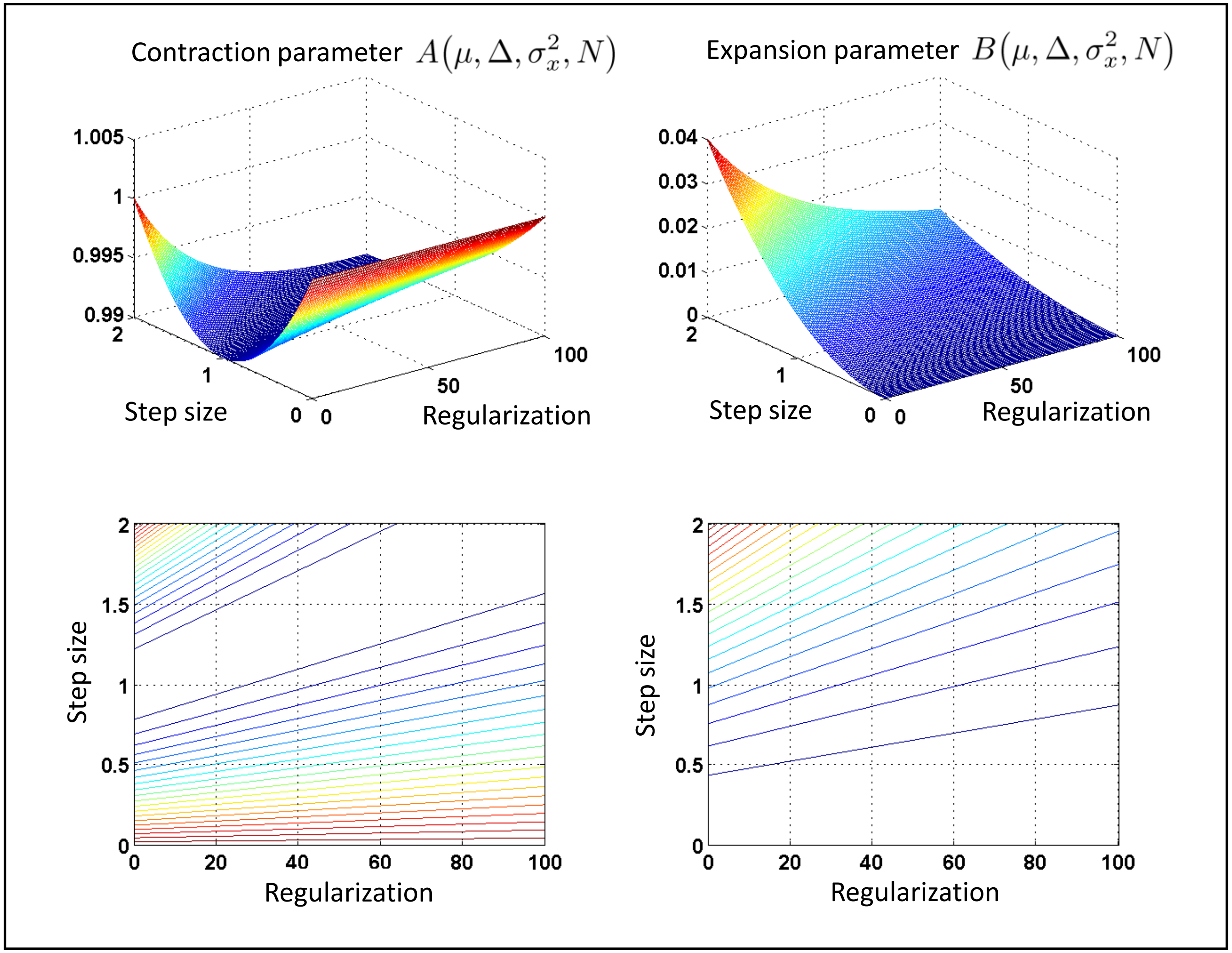
- Range: $0 \leq B(\mu,\Delta,\sigma_x^2,N) \leq \infty$
- Desired: as small as possible
- Determines the robustness against distortions

Opposite to each other – a common solution (optimization) Has to be found!

Influence of the Control Parameters

Values for the contraction and expansion parameters for the conditions:

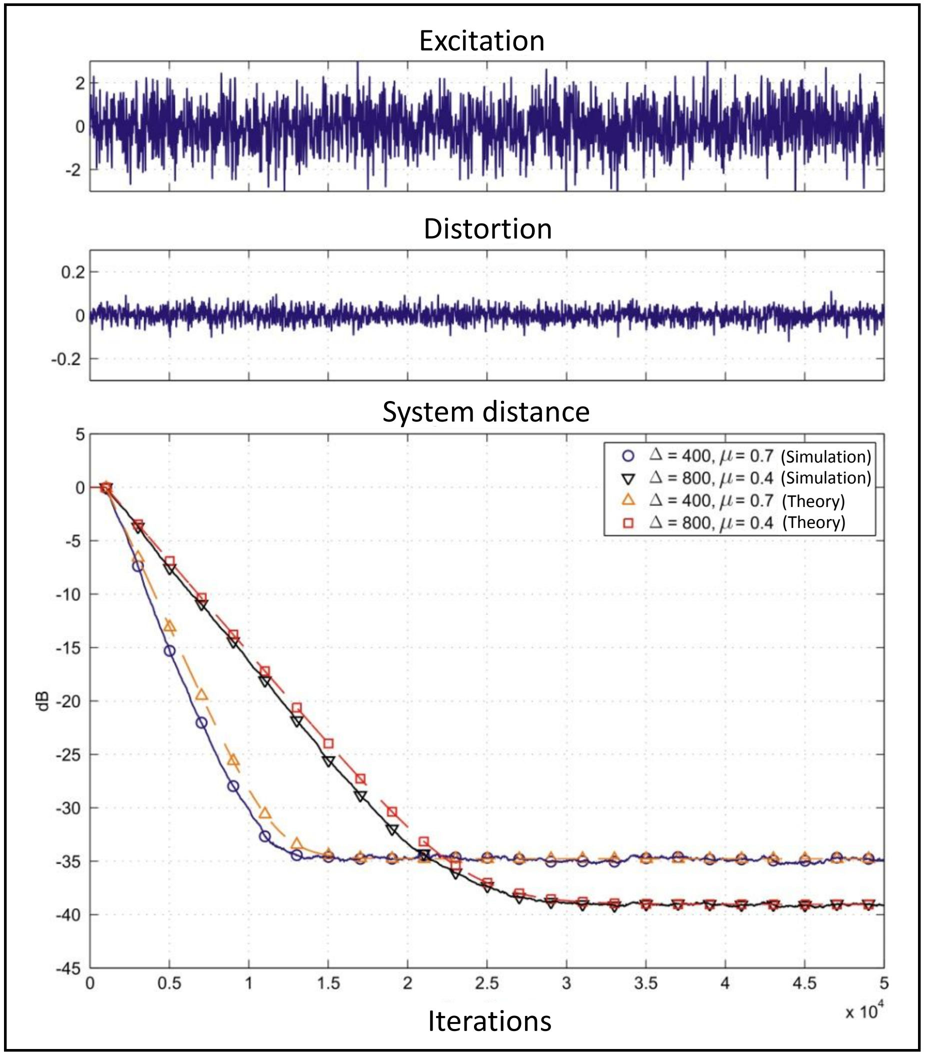
- $\sigma_x^2 = 1$
- $N = 100$



True and Prediction System Distance

Boundary conditions of the simulation:

- Excitation: white noise
- Distortion: white noise
- SNR: 30 dB



Maximum Convergence Speed – Part 1

For the special case without any distortions

$$n(n) \;=\; 0$$

and with optimal control parameters for that case

$$\mu(n) \;=\; 1, \;\Delta \;=\; 0$$

we get

$$\begin{aligned} \mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n+1)\|^2\right\} \;&\approx\; \left(1 + \frac{\mu^2\,N\,\sigma_x^4}{(N\,\sigma_x^2+\Delta)^2} - \frac{2\,\mu\,\sigma_x^2}{N\,\sigma_x^2+\Delta}\right)\,\mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n)\|^2\right\} \\ &=\; \left(1 - \frac{1}{N}\right)\,\mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n)\|^2\right\}. \end{aligned}$$

Meaning that the average system distance can be reduced per adaptation step by a factor of $1 - 1/N$. As a result adaptive filters with a lower amount of coefficients converge faster than long adaptive filters.

Maximum Convergence Speed – Part 2

If we want to know how long it takes to improve the filter convergence by 10 dB, we can make the following ansatz:

$$\mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n+n_{10\mathrm{dB}})\|^2\right\} = \frac{1}{10}\,\mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n)\|^2\right\} \approx \left(1-\frac{1}{N}\right)^{n_{10\mathrm{dB}}} \mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n)\|^2\right\}.$$

As on the previous slide we assumed an undisturbed adaptation process. By applying the natural logarithm we obtain

$$-\ln 10 \approx n_{10\,\mathrm{dB}} \ln\left(1-\frac{1}{N}\right).$$

By using the following approximations

$$\ln(1+x) \approx x \quad \text{for} \quad |x|<1, \qquad \text{and} \quad \ln 10 \approx 2,$$

we get

$$-2 \approx n_{10\,\mathrm{dB}}\left(-\frac{1}{N}\right)$$

$$\boxed{n_{10\,\mathrm{dB}} \approx 2N.}$$

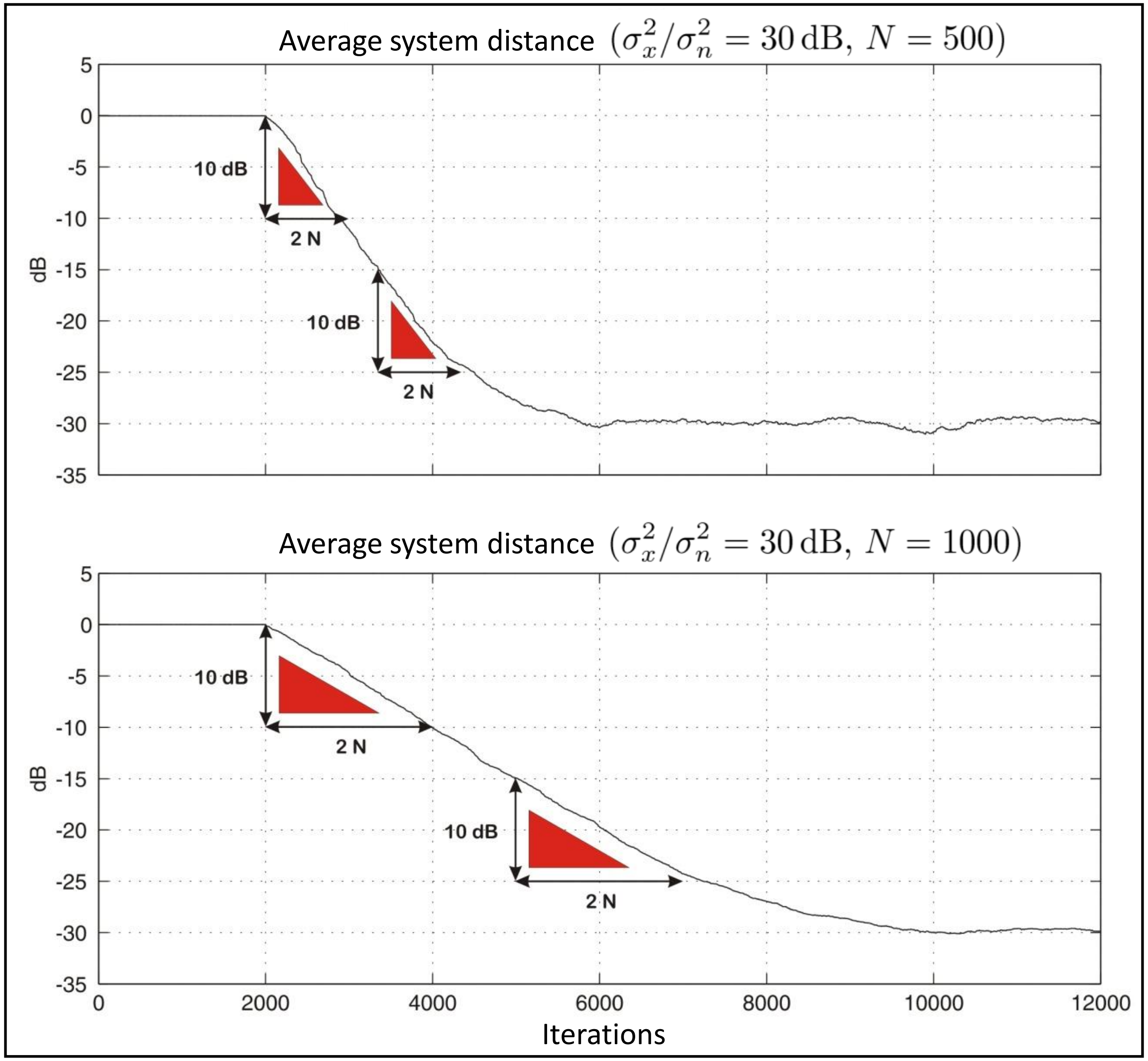
This means: At maximum speed of convergence it takes about 2N iterations until the average system distance is reduced by 10 dB.

Christian-Albrechts-Universität zu Kiel

The „10 dB per 2N“ Rule

Boundary conditions of the simulation:

- Excitation: white noise
- Distortion: white noise
- SNR: 30 dB
- Step size: 1
- Different filter lengths (500 and 1000)



Prediction of the Steady-State Convergence – Part 1

Recursion of the average system distance:

$$\begin{aligned}
\mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n)\|^2\right\} &\approx A(...)\,\mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n-1)\|^2\right\} + B(...)\,\sigma_n^2 \\
&\approx A(...)\left(A(...)\,\mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n-2)\|^2\right\} + B(...)\,\sigma_n^2\right) + B(...)\,\sigma_n^2 \\
&\ \ \vdots \\
&\approx A^n(...)\,\mathrm{E}\left\{\|\boldsymbol{h}_\Delta(0)\|^2\right\} + B(...)\,\sigma_n^2 \sum_{i=0}^{n-1} A^i(...)
\end{aligned}$$

For $n \to \infty$ and appropriately chosen control parameters we obtain:

$$\begin{aligned}
\lim_{n\to\infty} A^n(...)\,\mathrm{E}\left\{\|\boldsymbol{h}_\Delta(0)\|^2\right\} &= 0 \\
\lim_{n\to\infty} B(...)\,\sigma_n^2 \sum_{i=0}^{n-1} A^i(...) &= B(...)\,\sigma_n^2\,\frac{1}{1-A(...)}
\end{aligned}$$

Prediction of the Steady-State Convergence – Part 2

By inserting the results from the previous slide we obtain:

$$\lim_{n\to\infty} \mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n)\|^2\right\} \approx \frac{B(...)\,\sigma_n^2}{1-A(...)}$$

For the adaptation without regularization we get:

$$A(...) = 1 - \frac{\mu(2-\mu)}{N}$$

$$B(...) = \frac{\mu^2}{N\sigma_x^2}$$

Inserting these values leads to:

$$\boxed{\lim_{n\to\infty} \mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n)\|^2\right\} \approx \frac{\mu}{2-\mu}\,\frac{\sigma_n^2}{\sigma_x^2}}$$

Optimal Step Size – Motivation

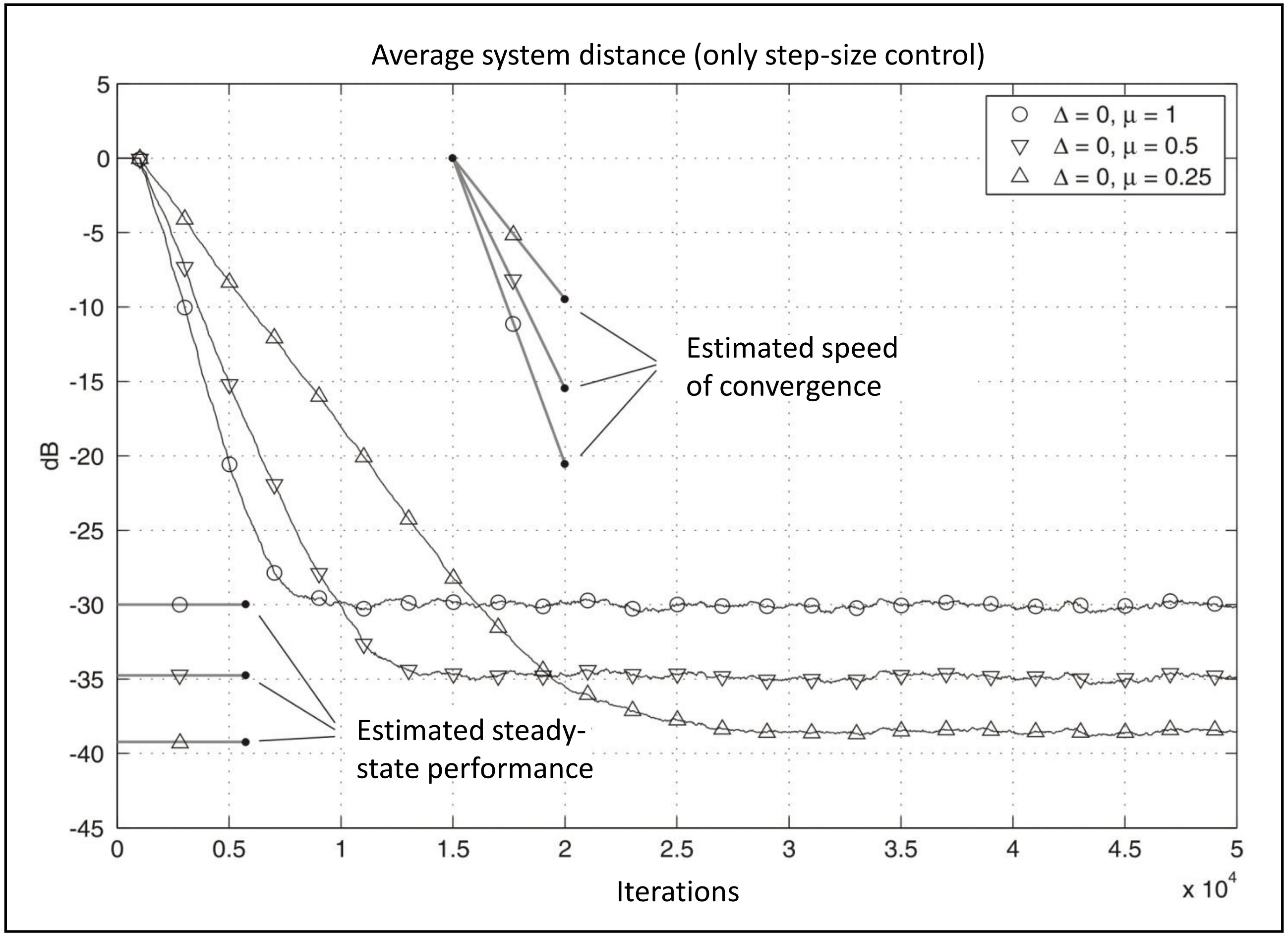
Remarks:

With a large step size one can achieve a fast initial convergence, but only a poor steady-state performance.

With a small step size a good steady-state performance can be obtained, but only a slow initial convergence.

Solution:

Utilization of a time-variant step-size.



Optimal Step Size – Derivation

… Derivation during the lecture …

Optimal Step Size – Example

Boundary conditions of the simulation:

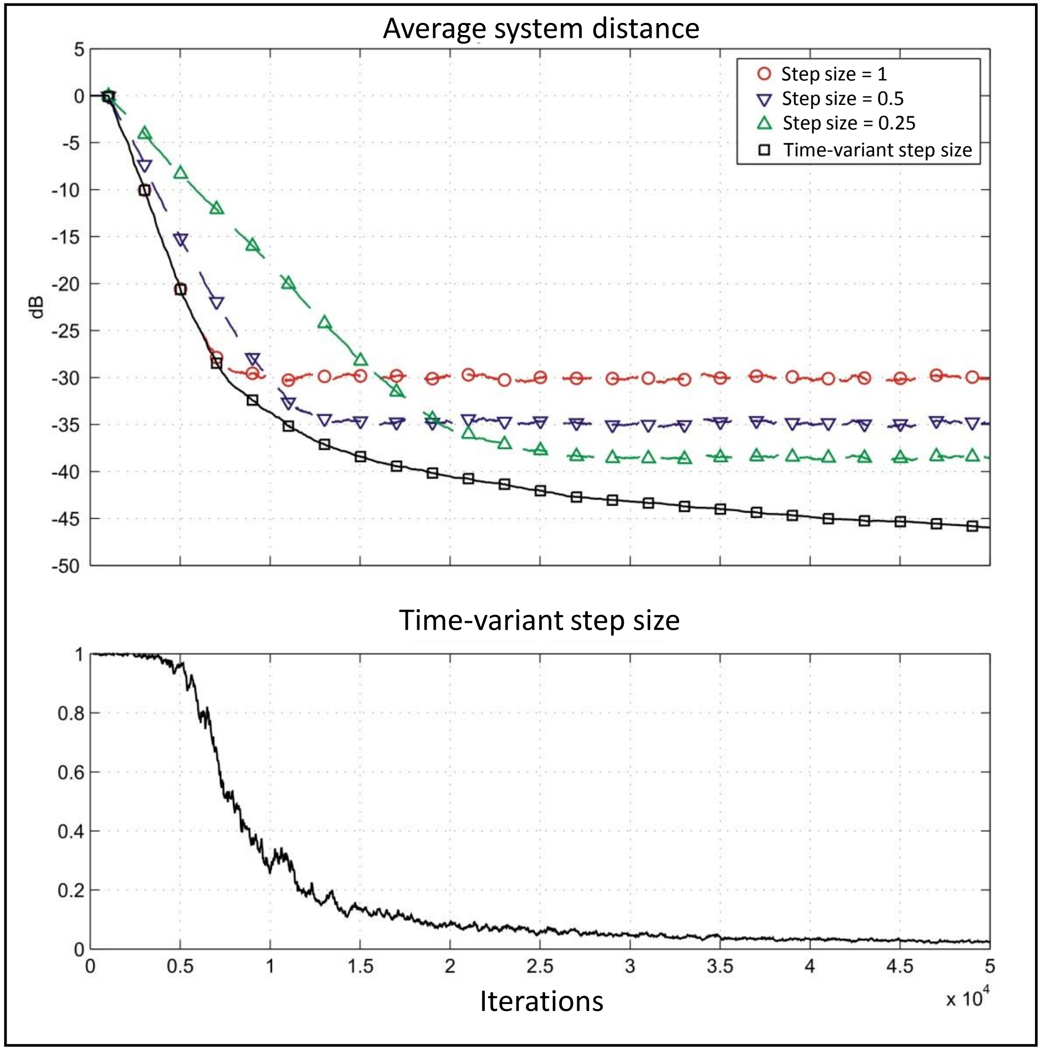
- Excitation: white noise
- Distortion: white noise
- SNR: 30 dB
- Filter length: 1000 coefficients

Computation of the step size:

$$P_e(n) = \beta \cdot P_e(n-1) + (1-\beta) \cdot e^2(n)$$
$$P_{e_u}(n) = \beta \cdot P_{e_u}(n-1) + (1-\beta) \cdot e_u^2(n)$$

with $\beta = 0.995$

$$\mu(n) = \frac{P_{e_u}(n)}{P_e(n)}$$



Estimation of the Optimal Step Size

Approximation for the optimal step size:

$$\mu_{\text{opt}}(n) \approx \frac{\mathrm{E}\Big\{e_{\mathrm{u}}^2(n)\Big\}}{\mathrm{E}\Big\{e^2(n)\Big\}} = \frac{\mathrm{E}\Big\{e_{\mathrm{u}}^2(n)\Big\}}{\mathrm{E}\Big\{e_{\mathrm{u}}^2(n)\Big\} + \mathrm{E}\Big\{n^2(n)\Big\}}$$

For white excitation we get:

$$\mu_{\text{opt}}(n) \approx \frac{\mathrm{E}\Big\{e_{\mathrm{u}}^2(n)\Big\}}{\mathrm{E}\Big\{e^2(n)\Big\}} = \frac{\mathrm{E}\Big\{x^2(n)\Big\} \cdot \big\|\boldsymbol{h}_{\Delta}(n)\big\|^2}{\mathrm{E}\Big\{e^2(n)\Big\}}$$

Ansatz:

Short-term power of the excitation signal

$$\widehat{\mu}_{\text{opt}}(n) = \frac{\overline{x^2(n)} \cdot b_{\mathrm{D}}(n)}{\overline{e^2(n)}}$$

Estimated system distance

Short-term power of the error signal

Christian-Albrechts-Universität zu Kiel

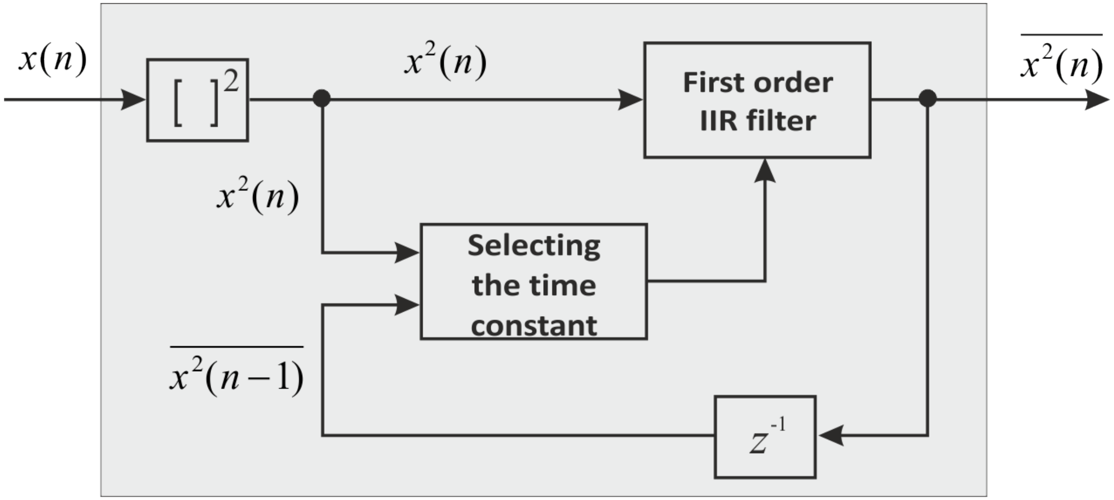
Estimation Procedures for the Optimal Step Size – Part 1

First order IIR smoothing with different time constants for rising and falling signal edges:

$$\overline{x^2(n)} = \begin{cases} \beta_\mathrm{r}\, x^2(n) + (1-\beta_\mathrm{r})\, \overline{x^2(n-1)}, & \text{if} \quad x^2(n) > \overline{x^2(n-1)}, \\ \beta_\mathrm{f}\, x^2(n) + (1-\beta_\mathrm{f})\, \overline{x^2(n-1)}, & \text{else.} \end{cases}$$

Different time constants are used to achieve smoothing on one hand but also being able to follow sudden signal increments quickly on the other hand.

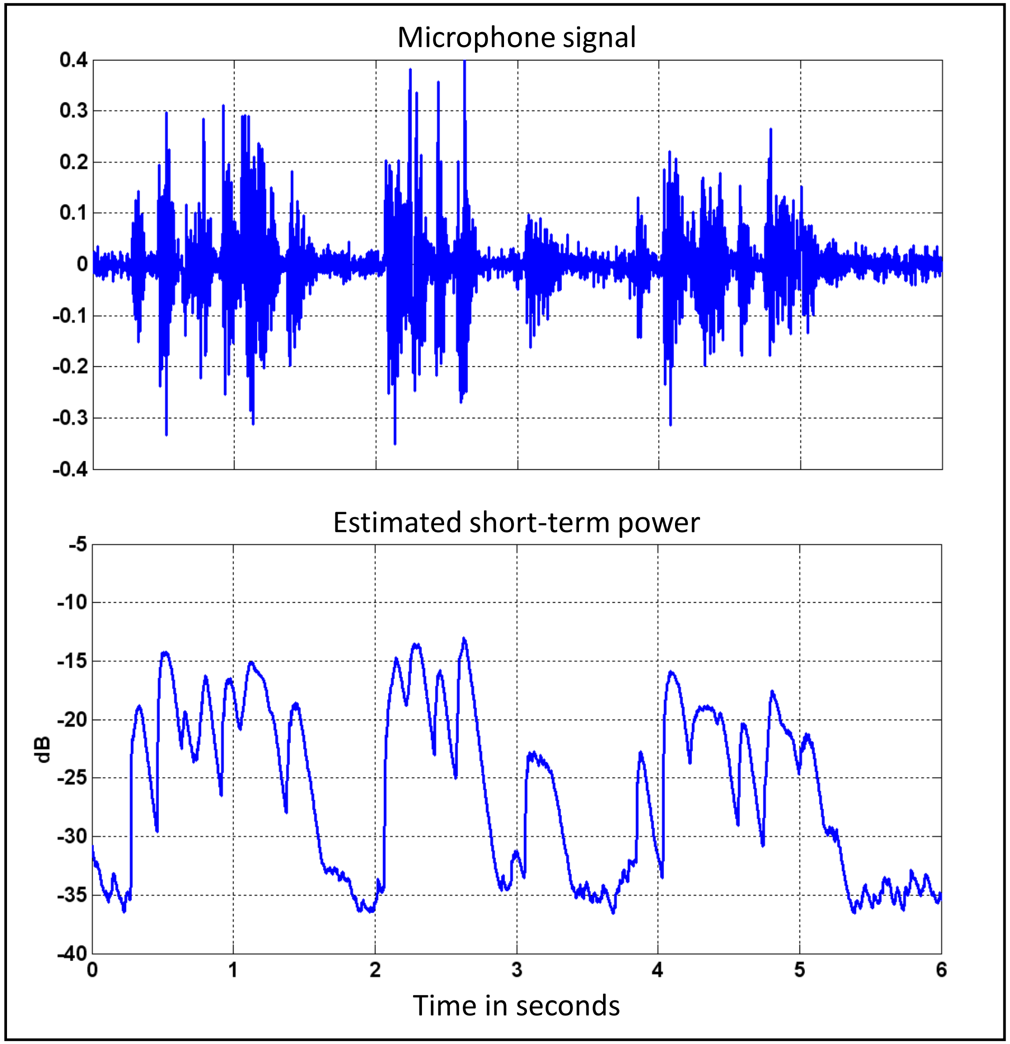
Basic structure:



Estimation Procedures for the Optimal Step Size – Part 2

Boundary conditions of the simulation:

- Excitation: speech
- SNR: about 20 dB
- $\beta_{\mathrm{r}} = 0.007, \ \beta_{\mathrm{f}} = 0.002$
- Sample rate: 8 kHz



Estimation Procedures for the Optimal Step Size – Part 3

Estimating the system distance:

$$\left\|\boldsymbol{h}_\Delta(n)\right\|^2 = \left\|\boldsymbol{h} - \widehat{\boldsymbol{h}}(n)\right\|^2 = \sum_{i=0}^{N-1} \left(h_i - \widehat{h}_i(n)\right)^2$$

Problem:

The coefficients h_i are not known.

Solution:

We extend the system by an artificial delay of N_D samples. For that part of the impulse response we have

$$h_i = 0, \qquad\qquad i \in \{0,\,1,\,...,\,N_\mathrm{D}-1\}.$$

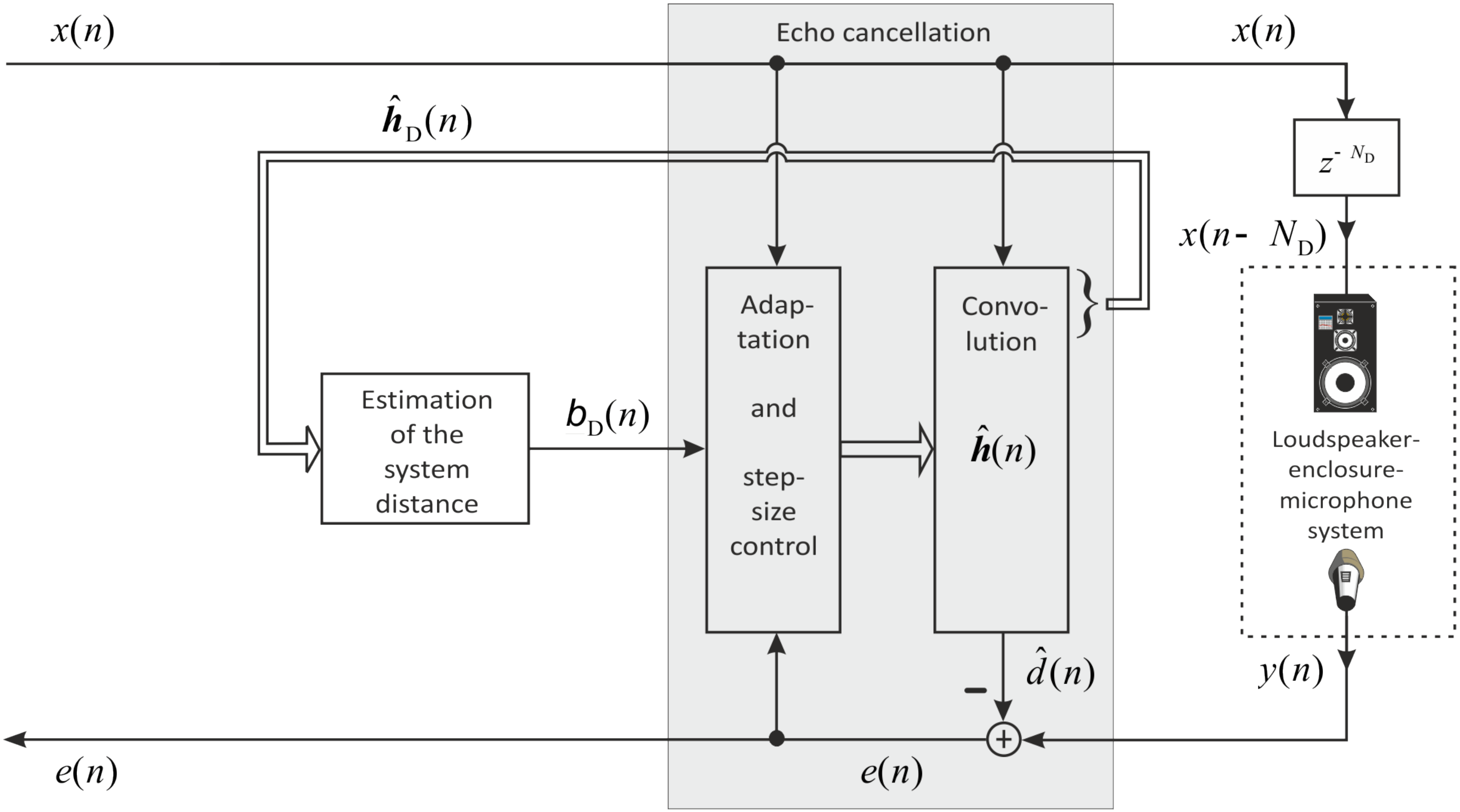
With these so-called ***delay coefficients*** we can extrapolate the system distance:

$$b_\mathrm{D}(n) = \frac{N}{N_\mathrm{D}} \sum_{i=0}^{N_\mathrm{D}-1} \widehat{h}_i^2(n).$$

Christian-Albrechts-Universität zu Kiel

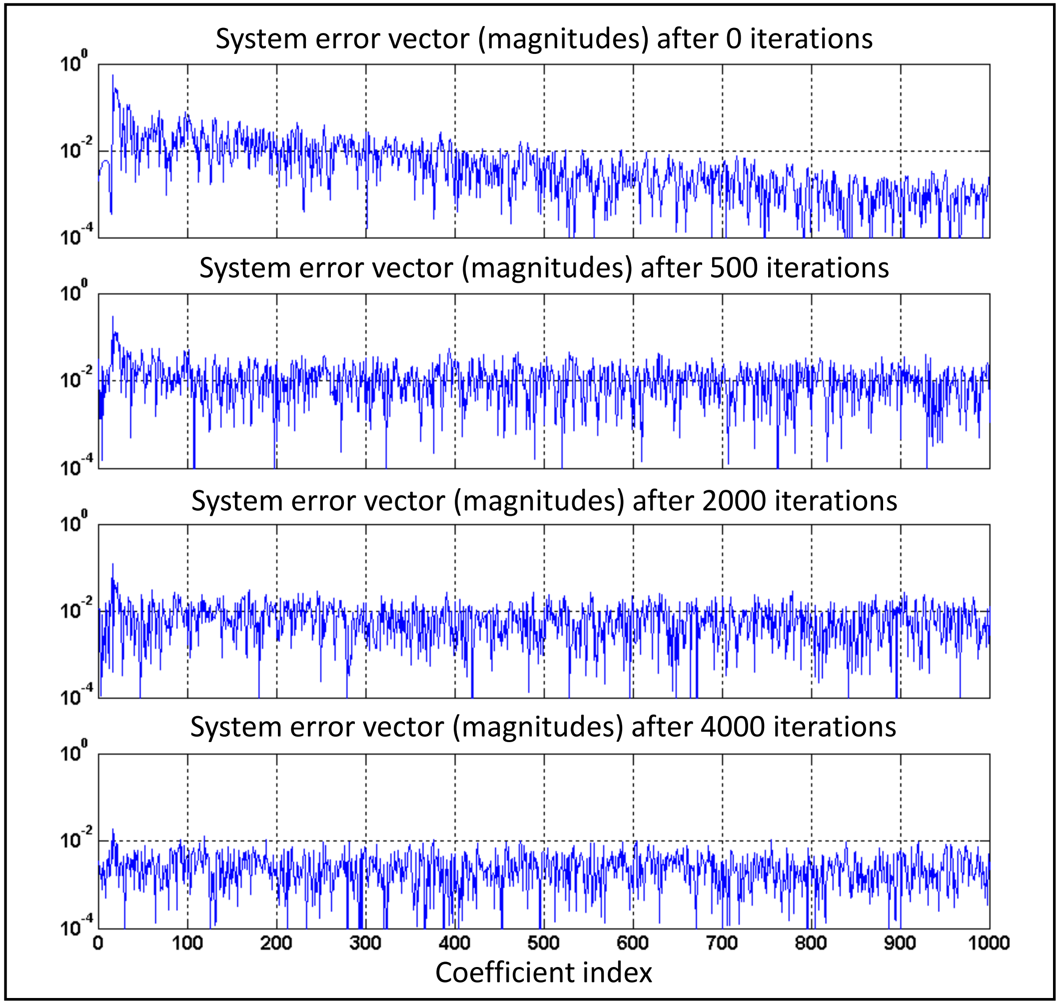
Estimation Procedures for the Optimal Step Size – Part 4

Structure of the system distance estimation:



Estimation Procedures for the Optimal Step Size – Part 5

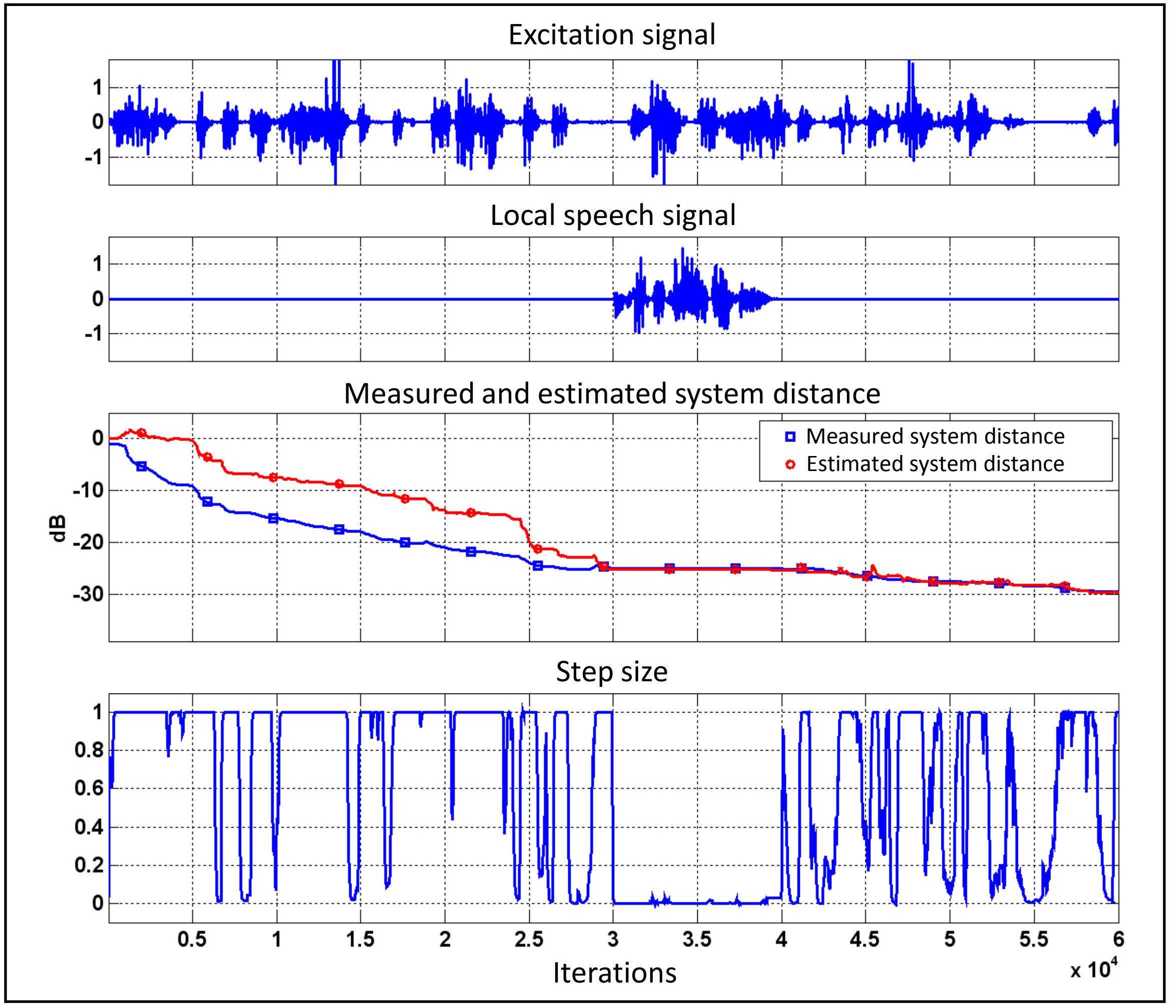
„Error spreading property" of the NLMS algorithm:



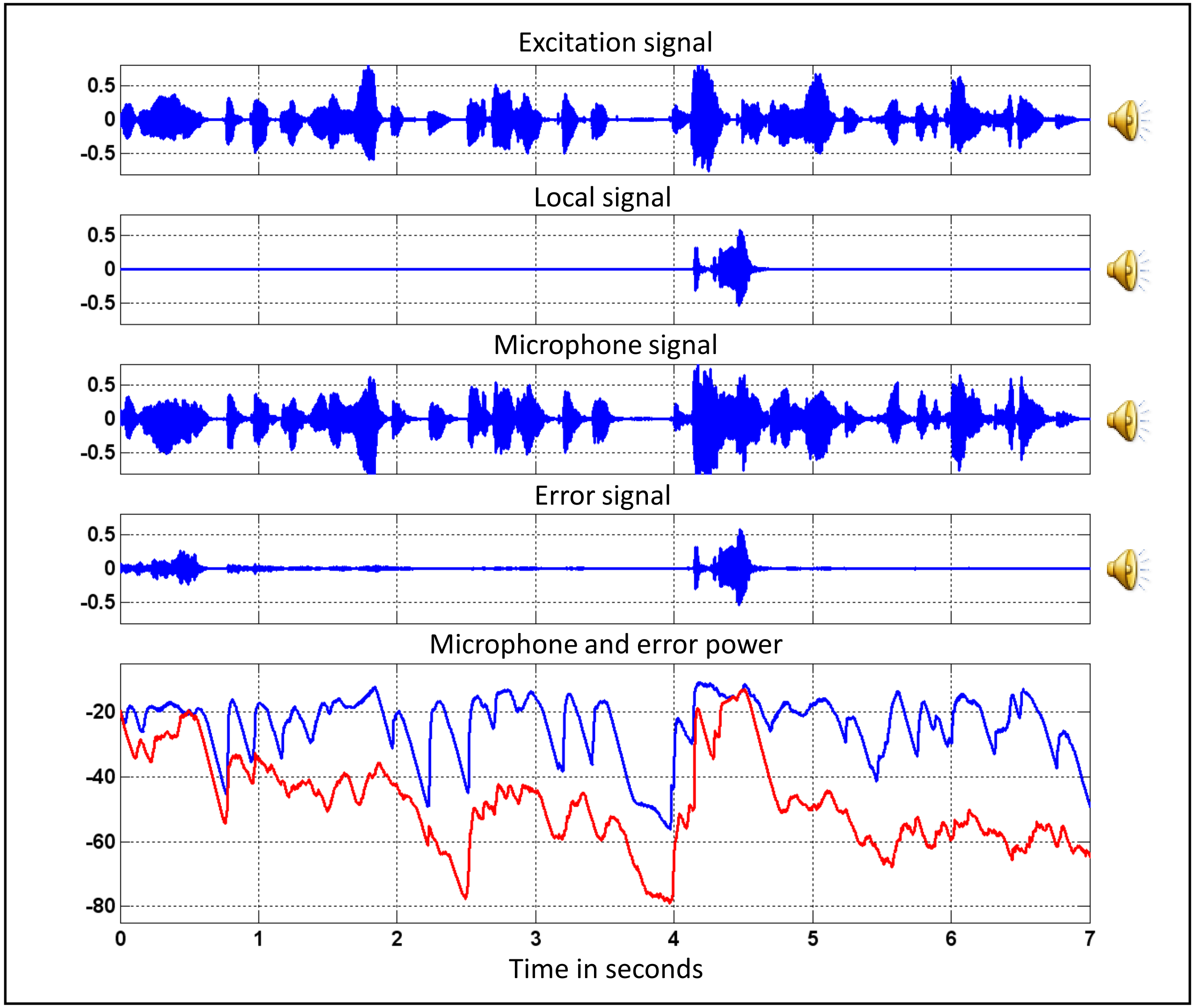
Estimation Procedures for the Optimal Step Size – Part 6

Boundary conditions of the simulation:

- Excitation: speech
- Distortion: speech
- SNR during single talk: 30 dB
- Filter length: 1000 coefficients



Convergence Examples – Part 5

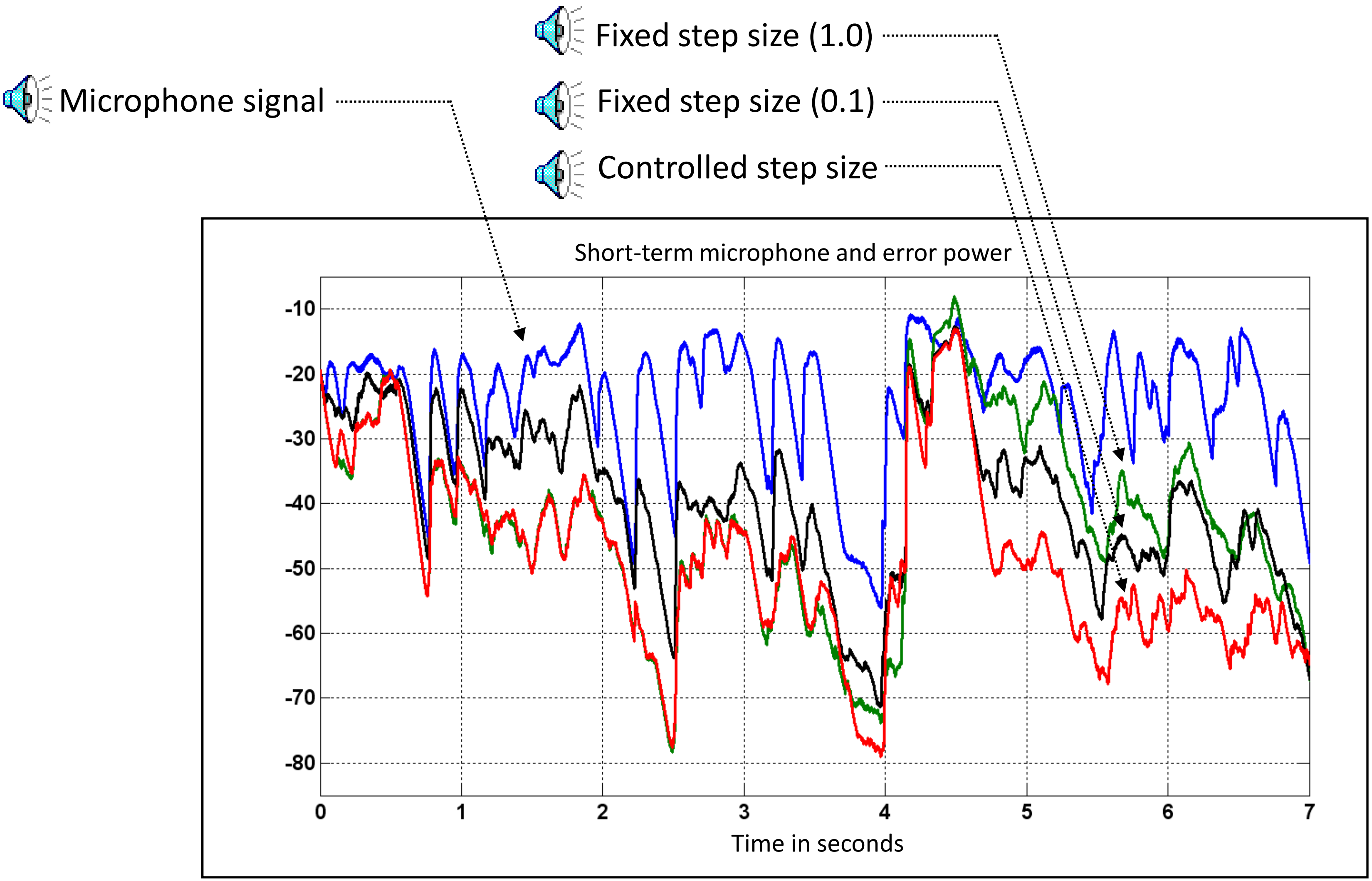


For Comparison:

- Fixed step size 1.0
- Fixed step size 0.1

Christian-Albrechts-Universität zu Kiel

Convergence Examples – Part 6



Summary and Outlook

This week:

- Introduction and Motivation
- Prediction of the System Distance
- Optimum Control Parameters
- Estimation Schemes
- Examples

Next week:

- Reducing the Computational Complexity of Adaptive Filters