Acoustic Echo Cancellation in the Presence of Continuous Double-Talk

Toon van Waterschoot*, Geert Rombouts*, Kris Struyve† and Marc Moonen*

SISTA Seminar 14/04/2005

*Katholieke Universiteit Leuven,
ESAT-SCD,
Kasteelpark Arenberg 10,
B-3001 Leuven, Belgium,
http://www.esat.kuleuven.ac.be/scd
toon.vanwaterschoot@esat.kuleuven.ac.be

†Televic N.V.,
L. Bekaertlaan 1,
B-8870 Izegem, Belgium,
http://www.televic.com/
k.struyve@televic.com



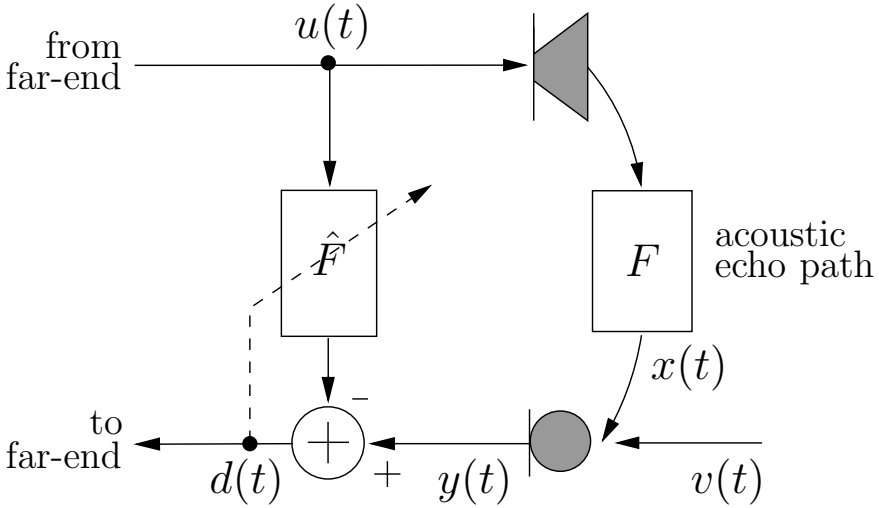


Overview

- Introduction:
  - Acoustic Echo Cancellation
  - Application: Teleconferencing
  - Application: Automatic gain adjustment
- Robust identification in continuous double-talk:
  - least squares estimator (LS)
  - minimum variance estimator (BLUE)
- Robust adaptive filtering:
  - Two-channel adaptive filtering algorithm (2ch-AF)
  - Prediction error method based adaptive filtering algorithm (PEM-AF)
  - Prediction error method based adaptive filtering algorithm with row operations (PEM-AFROW)
- Simulation results:
  - Gauss-Newton type adaptive algorithms
  - stochastic gradient type adaptive algorithms
- Conclusions

Introduction: Acoustic Echo Cancellation

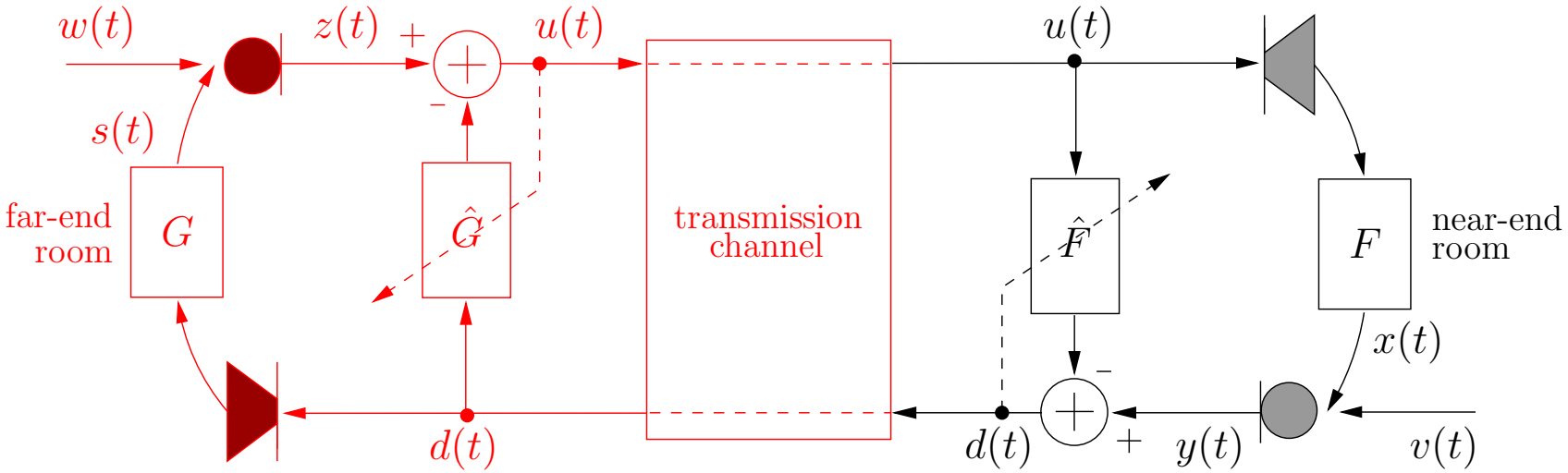
- Echo Problem:
  $y(t) = F(q,t)u(t) + v(t)$
- Echo Canceller:
  $d(t) = y(t) - \hat{F}(q,t)u(t)$
- Double-Talk:
  simultaneous far-end and near-end activity



- data vector definition:

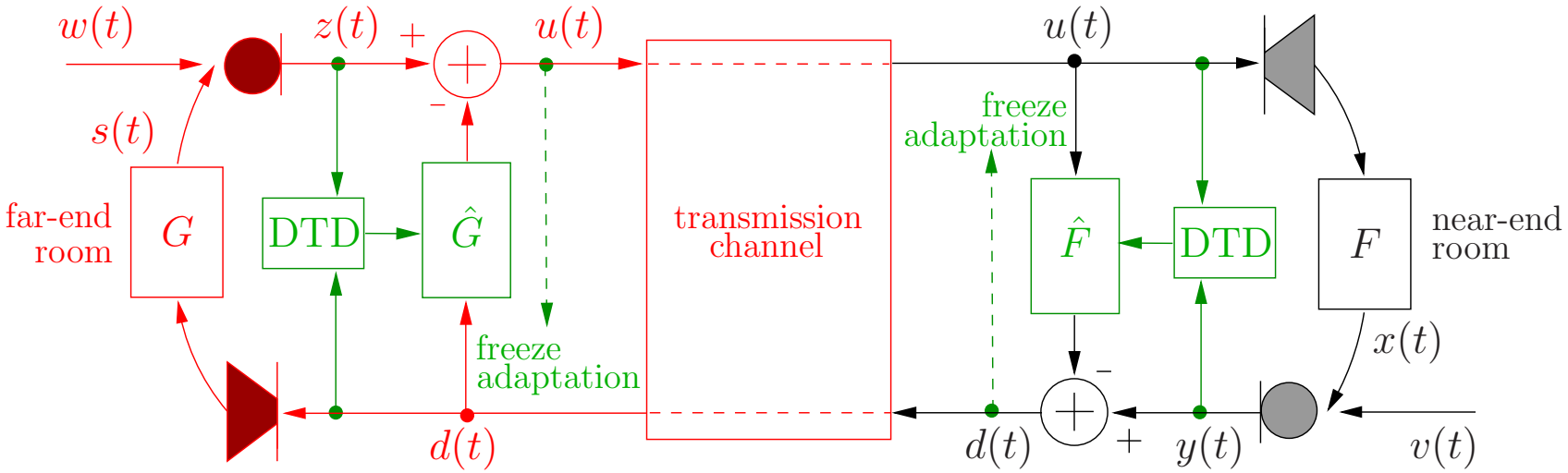
$$\underbrace{\begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(1) \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} u(t) & u(t-1) & \dots & u(t-n_F) \\ u(t-1) & u(t-2) & \dots & u(t-n_F-1) \\ \vdots & \vdots & \ddots & \vdots \\ u(1) & u(0) & \dots & u(1-n_F) \end{bmatrix}}_{\mathbf{U}} \cdot \underbrace{\begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_{n_F} \end{bmatrix}}_{\mathbf{f}} + \underbrace{\begin{bmatrix} v(t) \\ v(t-1) \\ \vdots \\ v(1) \end{bmatrix}}_{\mathbf{v}}$$

Introduction: Teleconferencing application (1)



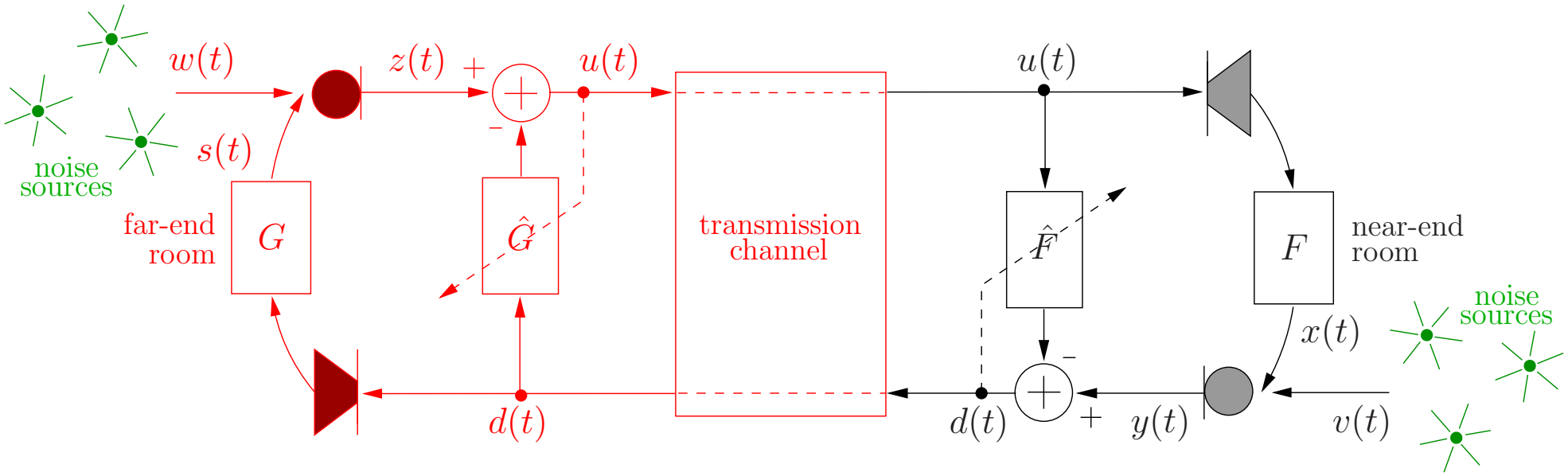
- Normal telephone conversation: 20% double-talk [Sondhi and Berkley, 1980]

Introduction: Teleconferencing application (2)



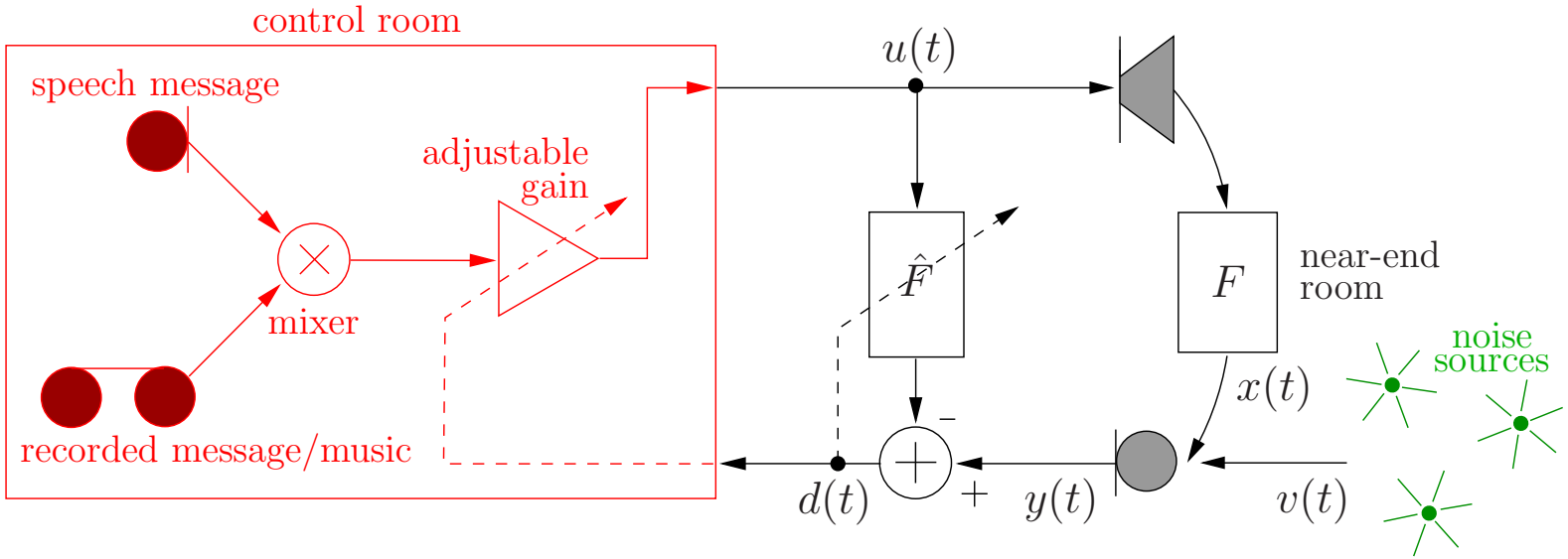
- Normal telephone conversation: 20% double-talk [Sondhi and Berkley, 1980]
  Tackled by freezing adaptation on double-talk detection (DTD)

Introduction: Teleconferencing application (3)



- Normal telephone conversation: 20% double-talk [Sondhi and Berkley, 1980]
  Tackled by freezing adaptation on double-talk detection (DTD)
- Background noise: 100% (continuous) double-talk
  Cannot be solved with a double-talk detector, only by adding robustness to the adaptive algorithm

Introduction: Automatic gain adjustment application



- Goal: to adjust the broadcasting gain to the background noise power, such that a constant SNR is maintained at the reference microphone
- Background noise: 100% (continuous) double-talk
  Cannot be solved with a double-talk detector, only by adding robustness to the adaptive algorithm

Robust identification in continuous double-talk: LS estimator

- Traditional algorithms for AEC are based on the least squares (LS) estimator, which can be calculated from the system of equations

$$\underbrace{\begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(1) \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} u(t) & u(t-1) & \dots & u(t-n_F) \\ u(t-1) & u(t-2) & \dots & u(t-n_F-1) \\ \vdots & \vdots & \ddots & \vdots \\ u(1) & u(0) & \dots & u(1-n_F) \end{bmatrix}}_{\mathbf{U}} \cdot \underbrace{\begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_{n_F} \end{bmatrix}}_{\mathbf{f}} + \underbrace{\begin{bmatrix} v(t) \\ v(t-1) \\ \vdots \\ v(1) \end{bmatrix}}_{\mathbf{v}}$$

  as: $\hat{\mathbf{f}}_{\mathbf{LS}} = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{y}$

- In a double-talk situation ($v(t) \neq 0$) the estimation of $\mathbf{f}$ is disturbed by the near-end signal. This leads to an increase of the estimator's variance, and as a consequence:
  - RLS: slower convergence, larger excess error in case of data windowing
  - APA: slower convergence or even divergence, larger excess error
  - NLMS: slower convergence or even divergence, larger excess error

Robust identification in continuous double-talk: BLUE

- A linear estimator of $\mathbf{f}$ having minimum variance in a double talk situation is given by the so-called best linear unbiased estimator (BLUE):

$$\hat{\mathbf{f}}_{\mathbf{BLUE}} = (\mathbf{U}^T\mathbf{R}^{-1}\mathbf{U})^{-1}\mathbf{U}^T\mathbf{R}^{-1}\mathbf{y},$$

  with $\mathbf{R}^{-1}$ the near-end signal correlation matrix:

$$\mathbf{R} \triangleq E\mathbf{v}\mathbf{v}^T = E\begin{bmatrix} v(t) \\ \vdots \\ v(1) \end{bmatrix}\begin{bmatrix} v(t) & \dots & v(1) \end{bmatrix}$$

- The near-end signal characteristics are obviously unknown and possibly time-varying. They need to be estimated concurrently with the room impulse response $\mathbf{f}$.

Robust adaptive filtering: prediction error method (PEM)

- The BLUE can be realized recursively by prefiltering the loudspeaker and microphone signals with the inverse near-end signal model and by including a weighting factor equal to the near-end excitation signal variance.
- If near-end signal $v(t)$ is modelled as an autoregressive sequence,

$$v(t) = \frac{1}{A(q,t)} e(t) \quad \text{with} \quad Ee(t)e(t-k) = \sigma_t^2 \delta(k),$$

the inverse near-end signal model and hence the prefilter equals $A(q,t)$, a finite impulse response (FIR) filter, and the optimal weight equals $\frac{1}{\sigma_t^2}$.

- This comes down to recursive minimization of the prediction error criterion

$$V_{PE}(t, A(q,t), F(q,t)) = \frac{1}{2N} \sum_{k=1}^{t} \frac{\lambda^{t-k}}{\hat{\sigma}_k^2} \big(A(q,t)[y(k) - F(q,t)u(k)]\big)^2.$$

with λ an appropriate exponential forgetting factor and $N = 1/(1-\lambda)$ the effective window length.

Robust adaptive filtering: 2ch-AF algorithm

- Minimization of:

$$
\begin{aligned}
V_{PE}(t, A(q,t), F(q,t)) &= \frac{1}{2N}\sum_{k=1}^{t}\frac{\lambda^{t-k}}{\hat{\sigma}_k^2}\big(\underbrace{A(q,t)}\, y(k) - \underbrace{A(q,t)F(q,t)}\, u(k)\big)^2 \\
&= \frac{1}{2N}\sum_{k=1}^{t}\frac{\lambda^{t-k}}{\hat{\sigma}_k^2}\big(A(q,t)y(k) + B(q,t)u(k)\big)^2 \\
&= \frac{1}{2N}\sum_{k=1}^{t}\frac{\lambda^{t-k}}{\hat{\sigma}_k^2}\big(y(k) + \boldsymbol{\phi_{2ch}}^T(k)\boldsymbol{\theta_{2ch}}(t)\big)^2
\end{aligned}
$$

with

$$
\begin{aligned}
\boldsymbol{\theta_{2ch}}(t) &\triangleq \begin{bmatrix} b_0(t) & \dots & b_{n_B}(t) & a_1(t) & \dots & a_{n_A}(t)\end{bmatrix}^T, \\
\boldsymbol{\phi_{2ch}}(k) &\triangleq \begin{bmatrix} u(k) & \dots & u(k-n_B) & y(k-1) & \dots & y(k-n_A)\end{bmatrix}^T
\end{aligned}
$$

leads to the two-channel adaptive filtering (2ch-AF) algorithm:

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{\boldsymbol{2ch}}(t) &= \hat{\boldsymbol{\theta}}_{\boldsymbol{2ch}}(t-1) - \frac{1}{\hat{\sigma}_t^2}\mathbf{R}^{-1}(t)\boldsymbol{\phi}_{\boldsymbol{2ch}}(t)\varepsilon(t, \hat{\boldsymbol{\theta}}_{\boldsymbol{2ch}}(t-1)), \\
\mathbf{R}(t) &= \lambda_{\boldsymbol{\theta}}\mathbf{R}(t-1) + \frac{1}{\hat{\sigma}_t^2}\boldsymbol{\phi}_{\boldsymbol{2ch}}(t)\boldsymbol{\phi}_{\boldsymbol{2ch}}{}^T(t), \\
\varepsilon(t, \hat{\boldsymbol{\theta}}_{\boldsymbol{2ch}}(t-1)) &= y(t) + \boldsymbol{\phi}_{\boldsymbol{2ch}}{}^T(t)\hat{\boldsymbol{\theta}}_{\boldsymbol{2ch}}(t-1).
\end{aligned}$$

- Complexity:
  - Gauss-Newton type: $\underbrace{4L_F^2 + 6L_F + 6}_{\text{WRLS complexity}} + \underbrace{16(L_A-1)L_F + 16L_A^2 - 20L_A + 4}_{\text{additional 2ch-AF complexity}} + \underbrace{L_F^2 - L_F}_{\text{deconvolution}}$
  - Stochastic gradient type: $\underbrace{7L_F + 1}_{\text{NLMS complexity}} + \underbrace{4L_A^2 + 5L_A + 1}_{\text{additional 2ch-AF complexity}} + \underbrace{L_F^2 - L_F}_{\text{deconvolution}}$
- Pro: convex cost function
  Contra: same data window for estimation of $F(q,t)$ and $A(q,t)$, deconvolution necessary

Robust adaptive filtering: PEM-AF algorithm

- Following the recursive prediction error method [Ljung and Söderström, 1986], minimization of

$$V_{PE}(t, A(q,t), F(q,t)) = \frac{1}{2N} \sum_{k=1}^{t} \frac{\lambda^{t-k}}{\hat{\sigma}_k^2} \big(A(q,t)[y(k) - F(q,t)u(k)]\big)^2$$

  with respect to both $\mathbf{f}(t)$ and $\mathbf{a}(t)$ leads to the recursion

$$\begin{bmatrix} \hat{\mathbf{f}}(t) \\ \hat{\mathbf{a}}(t) \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{f}}(t-1) \\ \hat{\mathbf{a}}(t-1) \end{bmatrix} + \frac{1}{\sigma_t^2} \mathbf{R}^{-1}(t) \begin{bmatrix} \boldsymbol{\psi_f}(t) \\ \boldsymbol{\psi_a}(t) \end{bmatrix} \varepsilon(t).$$

- The prediction error method based adaptive filtering (PEM-AF) algorithm is obtained by decoupling the adaptation of $\hat{\mathbf{f}}(t)$ and $\hat{\mathbf{a}}(t)$ by:
  - block-diagonalizing $\mathbf{R}^{-1}(t)$ [Spriet et al., 2004]:

$$\begin{bmatrix} \hat{\mathbf{f}}(t) \\ \hat{\mathbf{a}}(t) \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{f}}(t-1) \\ \hat{\mathbf{a}}(t-1) \end{bmatrix} + \frac{1}{\sigma_t^2} \begin{bmatrix} \mathbf{R_f}^{-1}(t) & \mathbf{0} \\ \mathbf{0} & \mathbf{R_a}^{-1}(t) \end{bmatrix} \begin{bmatrix} \boldsymbol{\psi_f}(t) \\ \boldsymbol{\psi_a}(t) \end{bmatrix} \varepsilon(t).$$

– introducing two different forgetting factors $\lambda_{\mathbf{f}}$ and $\lambda_{\mathbf{a}}$:

$$\begin{aligned}
\mathbf{R_f}(t) &= \lambda_{\mathbf{f}}\mathbf{R_f}(t-1) + \frac{1}{\hat{\sigma}_t^2}\boldsymbol{\psi_f}(t)\boldsymbol{\psi_f}^T(t), \\
\mathbf{R_a}(t) &= \lambda_{\mathbf{a}}\mathbf{R_a}(t-1) + \frac{1}{\hat{\sigma}_t^2}\boldsymbol{\psi_a}(t)\boldsymbol{\psi_a}^T(t).
\end{aligned}$$

– and approximating $\varepsilon(t)$, $\boldsymbol{\psi_f}(t)$ and $\boldsymbol{\psi_a}(t)$ as

$$\begin{aligned}
\varepsilon(t) &= \hat{A}(q,t-1)\xi(t), \\
\boldsymbol{\psi_f}(t) &= [\hat{A}(q,t-1)u(t) \ \ldots \ \hat{A}(q,t-n_F-1)u(t-n_F)]^T, \\
\boldsymbol{\psi_a}(t) &= \begin{bmatrix}\xi(t-1) & \ldots & \xi(t-n_A)\end{bmatrix}^T, \\
\xi(t) &\triangleq y(t) - \hat{F}(q,t-1)u(t).
\end{aligned}$$

- Complexity:
  - Gauss-Newton type: $\underbrace{4L_F^2 + 6L_F + 6}_{\text{WRLS complexity}} + \underbrace{2L_F + 4L_A^2 + 2L_A}_{\text{additional PEM-AF complexity}}$
  - Stochastic gradient type: $\underbrace{7L_F + 1}_{\text{NLMS complexity}} + \underbrace{2L_F + 4L_A^2 + 2L_A + 4}_{\text{additional PEM-AF complexity}}$
- Pro: decoupling of data windows for estimation of $F(q,t)$ and $A(q,t)$
  Contra: non-convex cost function (local minima), column and row operations (unless $A(q,t)$ is stationary on time scale of length L_F)

Robust adaptive filtering: PEM-AFROW algorithm

- Two-stage minimization of

$$V_{PE}(t, A(q,t), F(q,t)) = \frac{1}{2N} \sum_{k=1}^{t} \frac{\lambda^{t-k}}{\hat{\sigma}_k^2} \big(A(q,t)[y(k) - F(q,t)u(k)]\big)^2$$

yields the prediction error method based adaptive filtering algorithm with row operations (PEM-AFROW):

1. Estimate $A(q,t)$ and σ_t^2 with the Levinson-Durbin algorithm, on data frames with length M (e.g. $20ms$) and using a previous estimate of $F(q,t)$, and prefilter microphone and loudspeaker signals with $\hat{A}(q,t)$:

$$\begin{aligned} y_{\hat{\mathbf{a}}}(t) &= \hat{A}(q,t)y(t), \\ u_{\hat{\mathbf{a}}}(t) &= \hat{A}(q,t)u(t), \\ \boldsymbol{\psi_f}(t) &\triangleq \begin{bmatrix} u_{\hat{\mathbf{a}}}(t) & \dots & u_{\hat{\mathbf{a}}}(t-n_F) \end{bmatrix}^T. \end{aligned}$$

2. Estimate $\mathbf{f}(t)$ recursively:

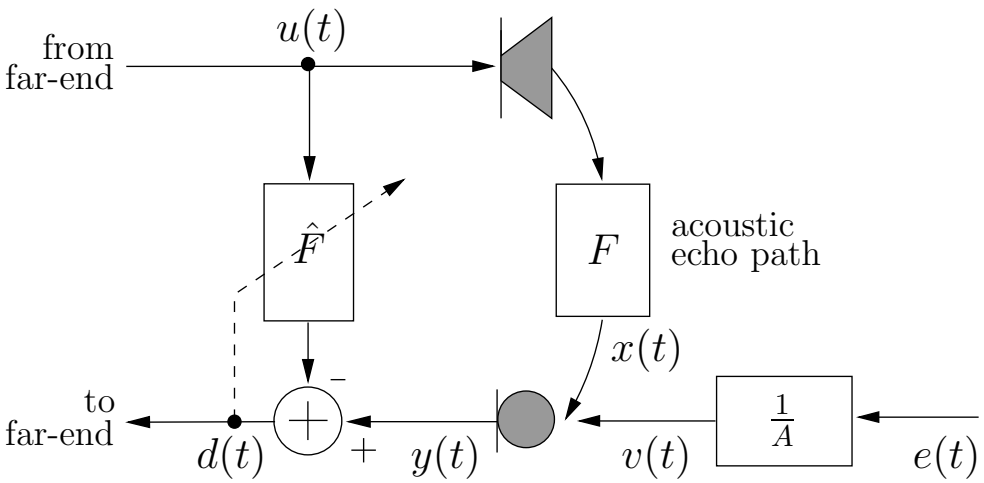
$$\begin{aligned}
\hat{\mathbf{f}}(t) &= \hat{\mathbf{f}}(t-1) + \frac{1}{\hat{\sigma}_t^2}\mathbf{R}^{-1}(t)\boldsymbol{\psi_f}(t)\varepsilon(t), \\
\mathbf{R}(t) &= \lambda\mathbf{R}(t-1) + \frac{1}{\hat{\sigma}_t^2}\boldsymbol{\psi_f}(t)\boldsymbol{\psi_f}^T(t), \\
\varepsilon(t,\hat{\mathbf{f}}(t-1)) &= y_A(t) - \boldsymbol{\psi_f}^T(t)\hat{\mathbf{f}}(t-1).
\end{aligned}$$

- Complexity:
  - Gauss-Newton type: $\underbrace{4L_F^2 + 6L_F + 6}_{\text{WRLS complexity}} + \underbrace{2L_F + 4L_A - 2 + \frac{L_F L_A - L_F + L_A^2 + L_A - 3}{M}}_{\text{additional PEM-AFROW complexity}}$
  - Stochastic gradient type: $\underbrace{7L_F + 1}_{\text{NLMS complexity}} + \underbrace{2L_F + 4L_A + 2 + \frac{L_F L_A - L_F + L_A^2 + L_A - 3}{M}}_{\text{additional PEM-AFROW complexity}}$

- Pro: decoupling of data windows for estimation of $F(q,t)$ and $A(q,t)$, only row operations
  Contra: non-convex cost function (local minima)

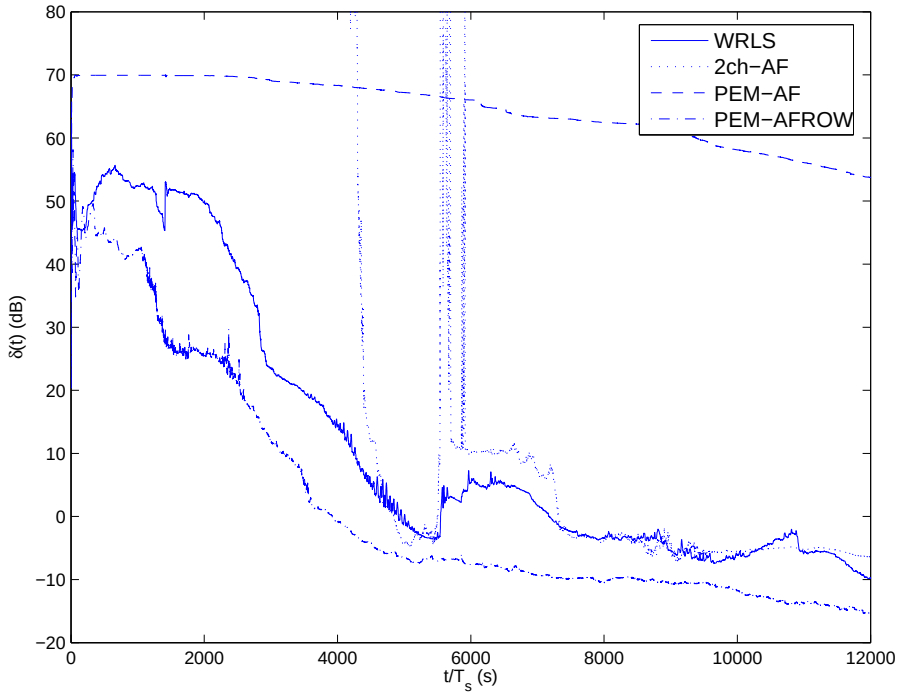
Simulation results (1)

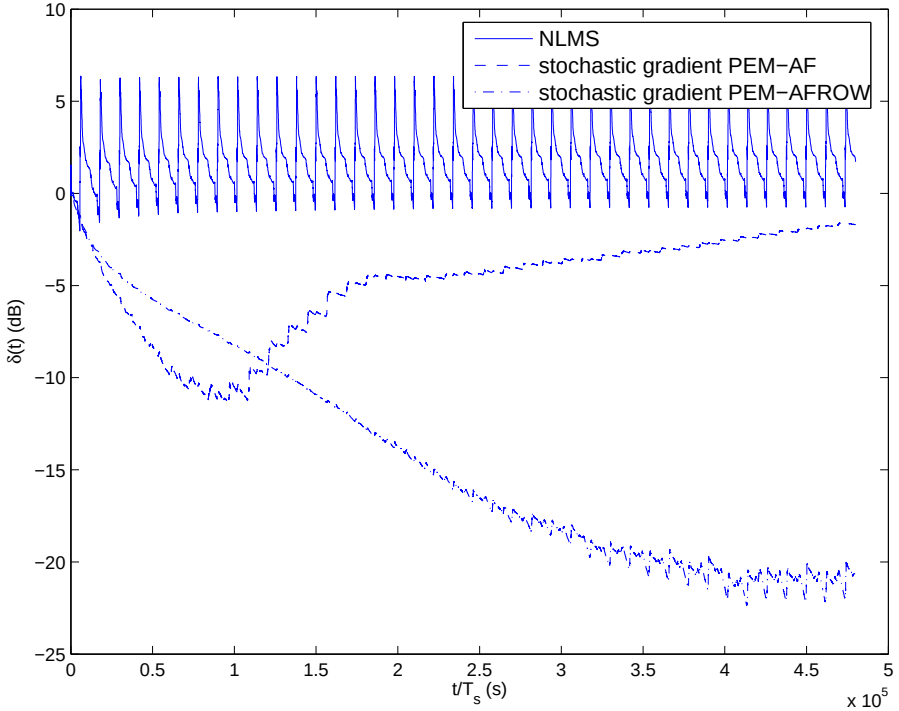
- Simulation parameters:
  - $f_s = 8kHz$
  - $n_F = 1000$
  - $n_A = 55$
  - $u(t) =$ male speech
  - $v(t) =$ female speech
  - echo-to-background ratio $\frac{\sum_{k=1}^{N} |x(k)|^2}{\sum_{k=1}^{N} |v(k)|^2} = 10dB$



- Performance measure: $\delta(t) = 20 \log_{10} \frac{\|\hat{\mathbf{f}}(t) - \mathbf{f}\|}{\|\mathbf{f}\|}$.
- The proposed algorithms have been compared in two flavours:
  - with Gauss-Newton update: WRLS, 2ch-AF, PEM-AF, PEM-AFROW
  - with stochastic gradient update: NLMS, PEM-AF, PEM-AFROW.

Simulation results (2)





Conclusions

- A continuous double-talk situation, which may occur e.g. in a noise teleconferencing environment or an automatic gain adjustment application, cannot be tackled using a double-talk detector.
- The robustness of an adaptive algorithm for acoustic echo cancellation may be improved by prefiltering the loudspeaker and microphone signals with the inverse near-end signal model and by including a weighting factor equal to the near-end excitation signal variance.
- Three algorithms for estimating the room impulse response concurrently with the near-end signal model were presented in a common framework: 2ch-AF, PEM-AF and PEM-AFROW.
- The 2ch-AF and PEM-AF algorithms rely on the assumption that the near-end signal characteristics do not change significantly on a time scale comparable to the room impulse response length. They do not perform better than a WRLS or NLMS algorithm.
- The PEM-AFROW algorithm which performs only row operations to the loudspeaker signal data matrix, shows an improved convergence compared to the other algorithms.