Acoustic Echo Cancellation in the Presence of Continuous Double-Talk

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

SPS-DARTS 2005, April 19-20, 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

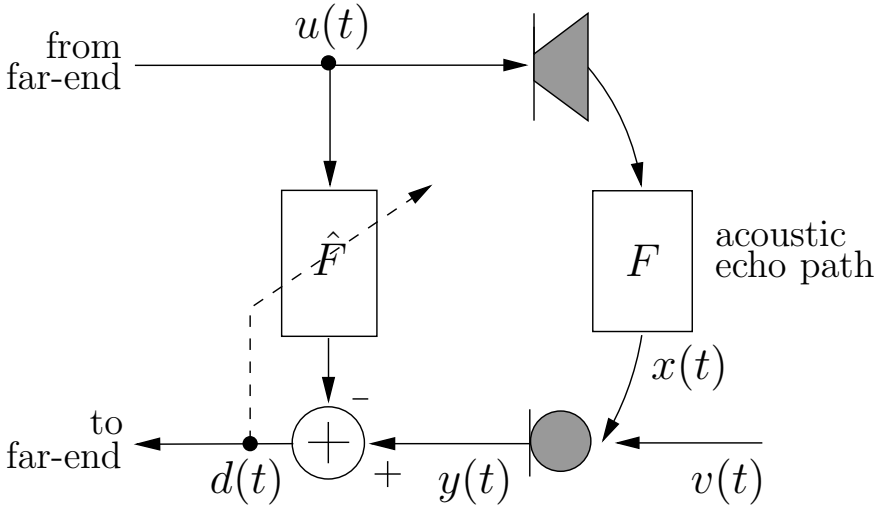
†Televic N.V.,
L. Bekaertlaan 1,
B-8870 Izegem, Belgium,
http://www.televic.com/
k.struyve@televic.com

Overview

- Introduction:
  - Acoustic Echo Cancellation
  - Application: Teleconferencing
  - Application: Automatic gain adjustment
  - Application: Feedback cancellation
- Robust identification in continuous double-talk:
  - least squares estimator (LS)
  - minimum variance estimator (BLUE)
- Robust adaptive filtering (from feedback cancellation):
  - 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

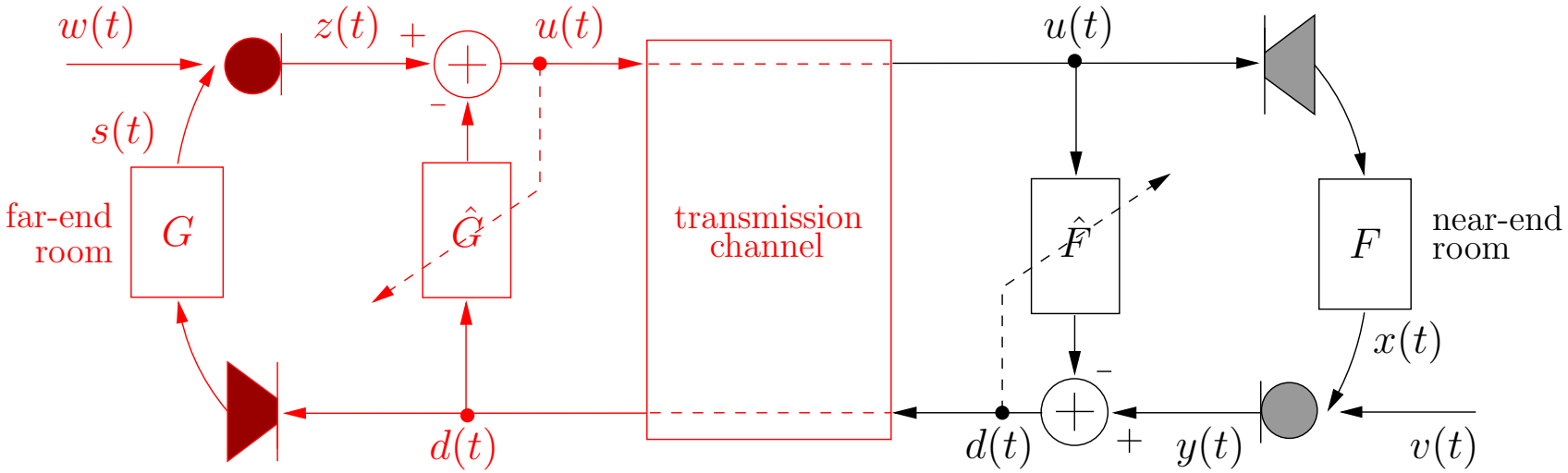
- Echo Problem:
  $y(t) = F(q,t)u(t) + v(t)$
- Echo Canceller:
  $d(t) = y(t) - \hat{F}(q,t)u(t)$
- Double-Talk:
  simultaneous far-end and near-end activity



- data vector definition:

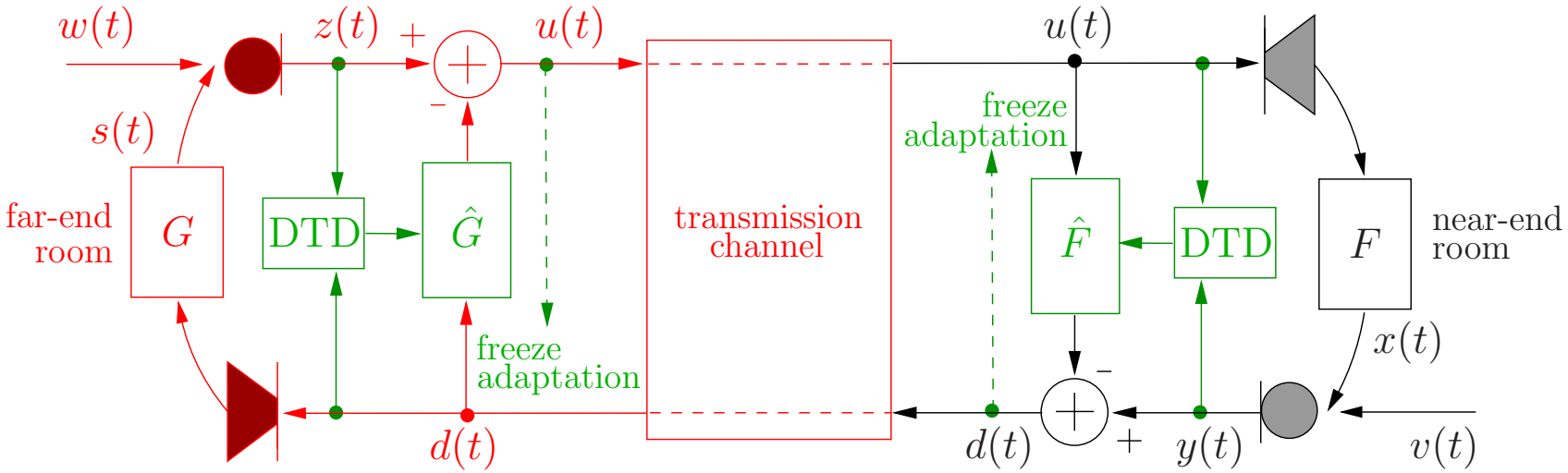
$$\underbrace{\begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(1) \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} u(t) & u(t-1) & \dots & u(t-n_F) \\ u(t-1) & u(t-2) & \dots & u(t-n_F-1) \\ \vdots & \vdots & \ddots & \vdots \\ u(1) & u(0) & \dots & u(1-n_F) \end{bmatrix}}_{\mathbf{U}} \cdot \underbrace{\begin{bmatrix}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

Introduction: Teleconferencing application (1)



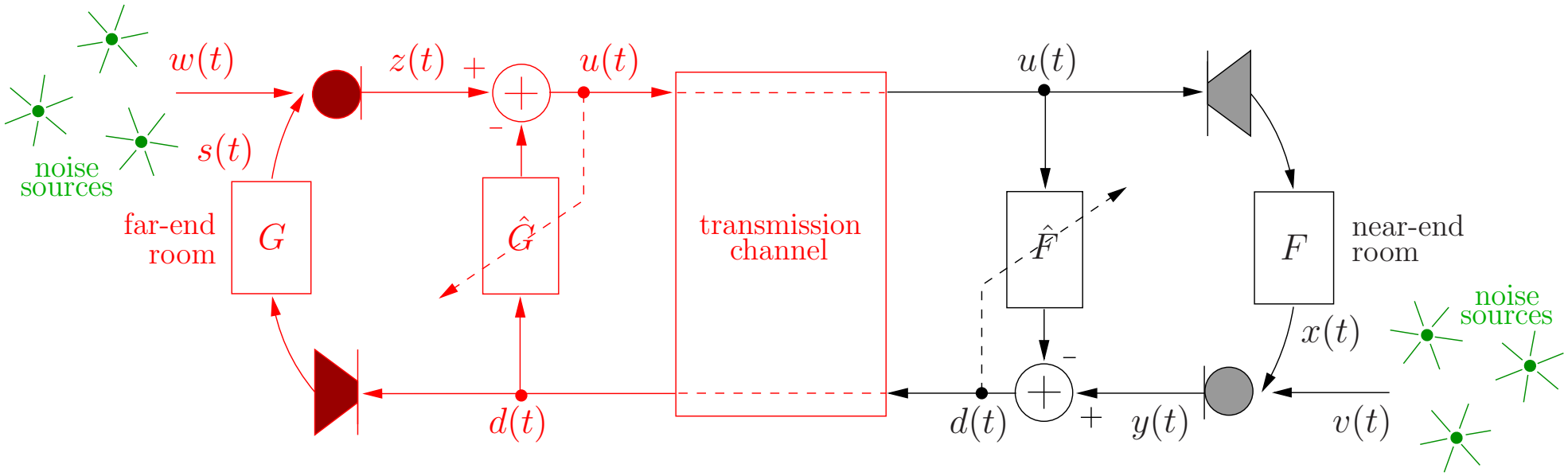
- Normal telephone conversation: 20% double-talk [Sondhi and Berkley, 1980]

Introduction: Teleconferencing application (2)



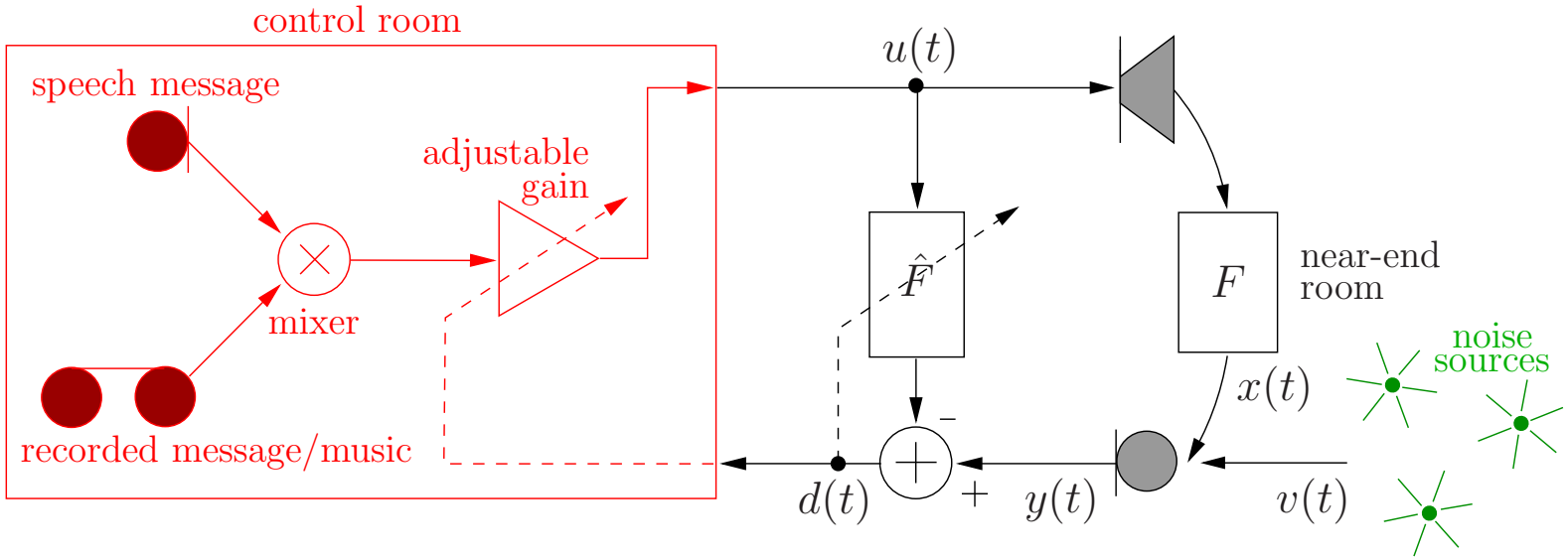
- Normal telephone conversation: 20% double-talk [Sondhi and Berkley, 1980]
  Tackled by freezing adaptation on double-talk detection (DTD)

Introduction: Teleconferencing application (3)



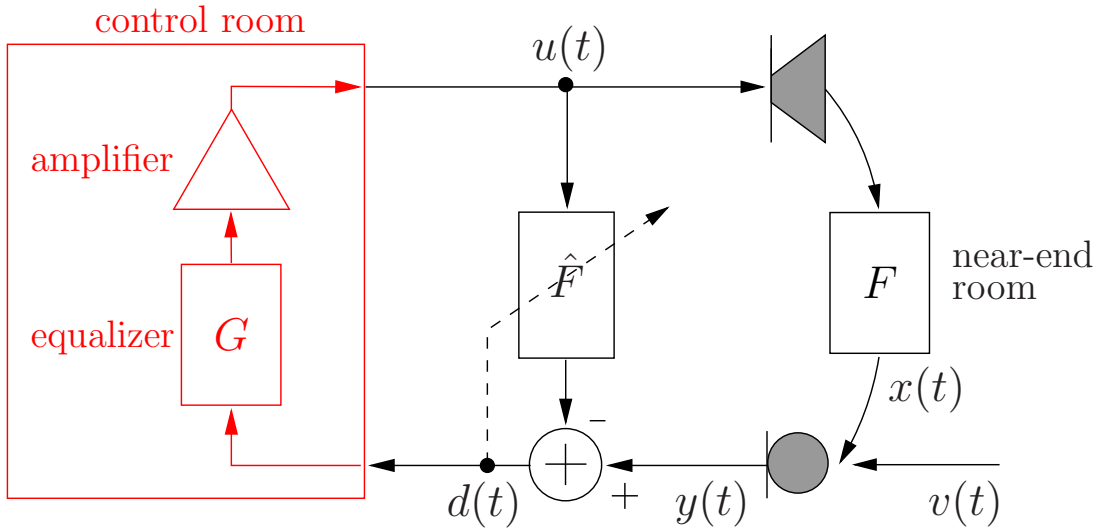
- Normal telephone conversation: 20% double-talk [Sondhi and Berkley, 1980]
  Tackled by freezing adaptation on double-talk detection (DTD)
- Background noise: 100% (continuous) double-talk
  Cannot be solved with a double-talk detector, only by adding robustness to the adaptive algorithm

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

Introduction: Feedback canceller application



- Far-end signal is an equalized and amplified version of near-end signal: 100% (continuous) double-talk
- Correlated double-talk situation: decorrelation needed to avoid biased feedback canceller

Robust identification in continuous double-talk: LS estimator (1)

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

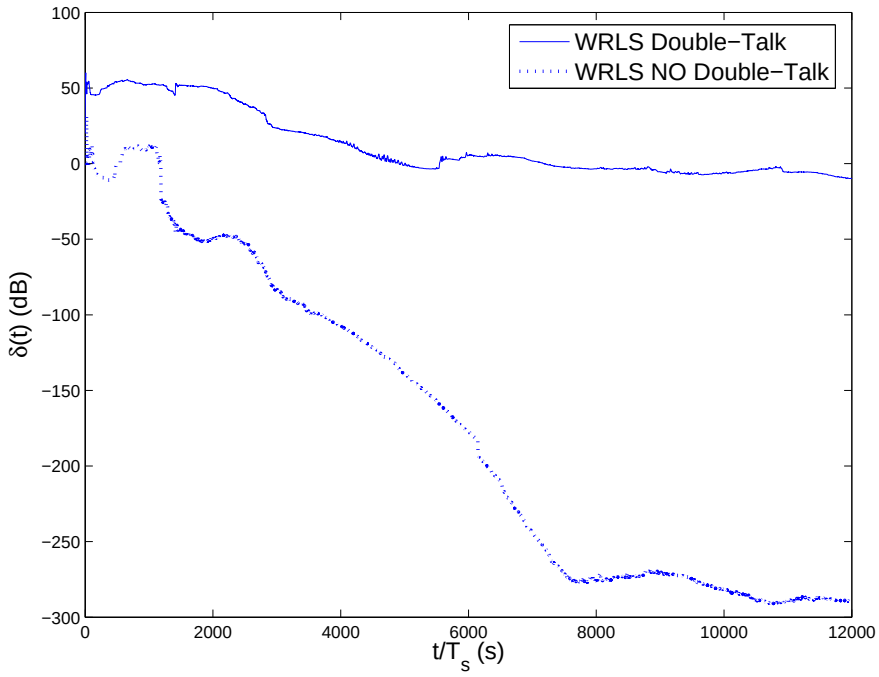
$$\underbrace{\begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(1) \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} u(t) & u(t-1) & \dots & u(t-n_F) \\ u(t-1) & u(t-2) & \dots & u(t-n_F-1) \\ \vdots & \vdots & \ddots & \vdots \\ u(1) & u(0) & \dots & u(1-n_F) \end{bmatrix}}_{\mathbf{U}} \cdot \underbrace{\begin{bmatrix}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

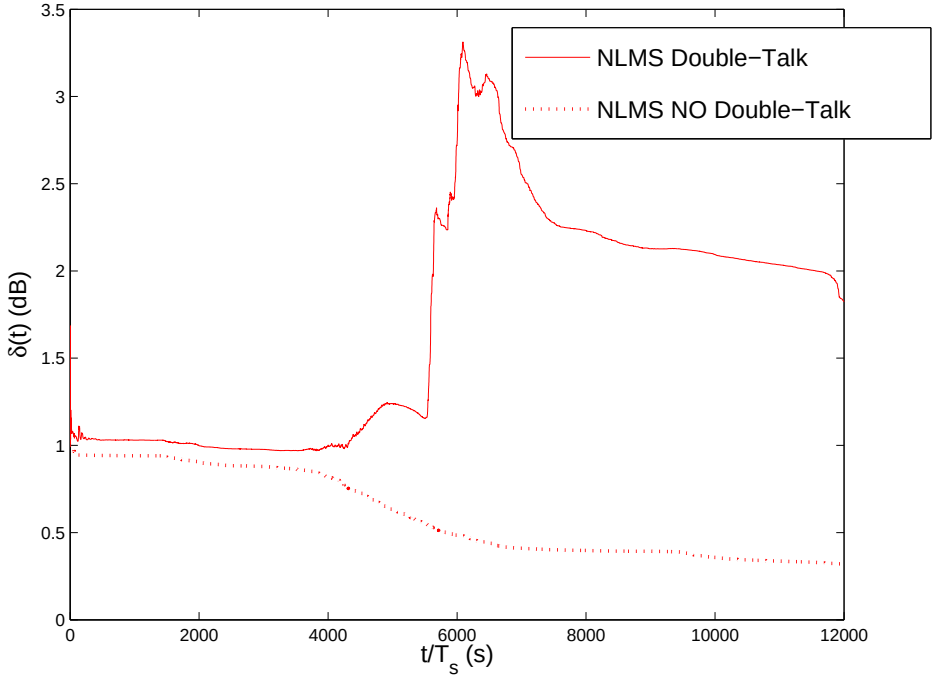
  as: $\hat{\mathbf{f}}_{\mathbf{LS}} = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{y}$

- 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: LS estimator (2)

Comparison of normalized distance $\delta(t) = 20\log_{10}\frac{\|\hat{\mathbf{f}}(t)-\mathbf{f}\|}{\|\mathbf{f}\|}$ between the true and estimated room impulse response for RLS and NLMS adaptive algorithms with and without double-talk.





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 $\mathbf{f}$.

- Note that for $\mathbf{R} = \mathbf{I}$ the BLUE is equivalent to the LS estimator. So the LS estimator is optimal only in case the near-end signal is drawn from a stationary white noise process.

Robust adaptive filtering: prediction error method (PEM) (1)

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

If we assume an autoregressive model for the near-end signal:

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

the near-end signal correlation matrix $\mathbf{R}^{-1}$ can be decomposed as:

$$\mathbf{R}^{-1} = \underbrace{\begin{bmatrix} A(q,t) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & A(q,1) \end{bmatrix}}_{\text{prefiltering}} \underbrace{\begin{bmatrix} \frac{1}{\sigma_t^2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sigma_{1-n_A}^2} \end{bmatrix}}_{\text{weighting}} \underbrace{\begin{bmatrix} A(q,t) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & A(q,1) \end{bmatrix}^T}_{\text{prefiltering}},$$

where $A(q,k)$ are finite impulse response (FIR) prefilters.

Robust adaptive filtering: prediction error method (PEM) (2)

- LS criterion:

$$V_{LS}(t, F(q,t)) = \frac{1}{2N}\sum_{k=1}^{t} \lambda^{t-k}\big(y(k) - F(q,t)u(k)\big)^2,$$

  WLS criterion = LS criterion + weighting:

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

  PE criterion = LS criterion + weighting + prefiltering:

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

- The prediction error criterion should be minimized w.r.t. both $F(q,t)$ and $A(q,t)$.

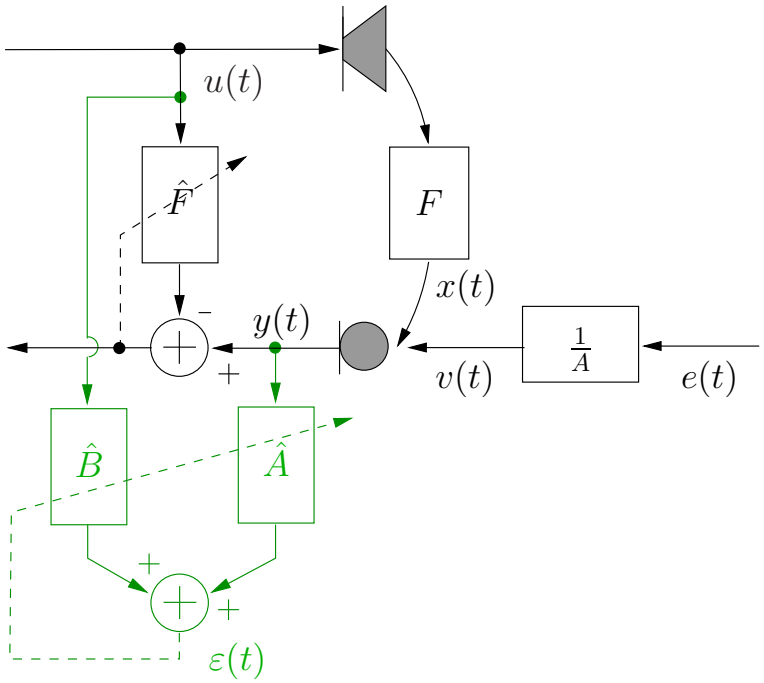
Robust AF Solution 1: 2ch-AF algorithm [Spriet et al., 2002]

- Linearization of the PE criterion leads to the two-channel adaptive filtering (2ch-AF) algorithm:

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

- 2ch-AF algorithm:
  - Pro: convex cost function (no local minima)
  - Contra: same data window for estimation of $F(q,t)$ and $A(q,t)$,

    deconvolution needed to obtain $F(q,t)$ from $B(q,t)$ (complexity)



Robust AF Solution 2 & 3: PEM based adaptive filtering

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

- RLS recursion for estimation of $F(q,t)$:

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

- Questions:
  1. How to calculate the prefiltered microphone signal $y_A(t)$?
     Exact: $y_A(t) = A(q,t)y(t) \quad \rightarrow$ Estimate: $y_A(t) = \hat{A}(q,t-1)y(t)$
  2. How to calculate the gradient vector $\boldsymbol{\psi_f}(t)$?
     Exact: $\boldsymbol{\psi_f}(t) = \big[A(q,t)u(t) \quad \ldots \quad A(q,t)u(t-n_F)\big]^T$
  3. How to estimate $A(q,t)$?

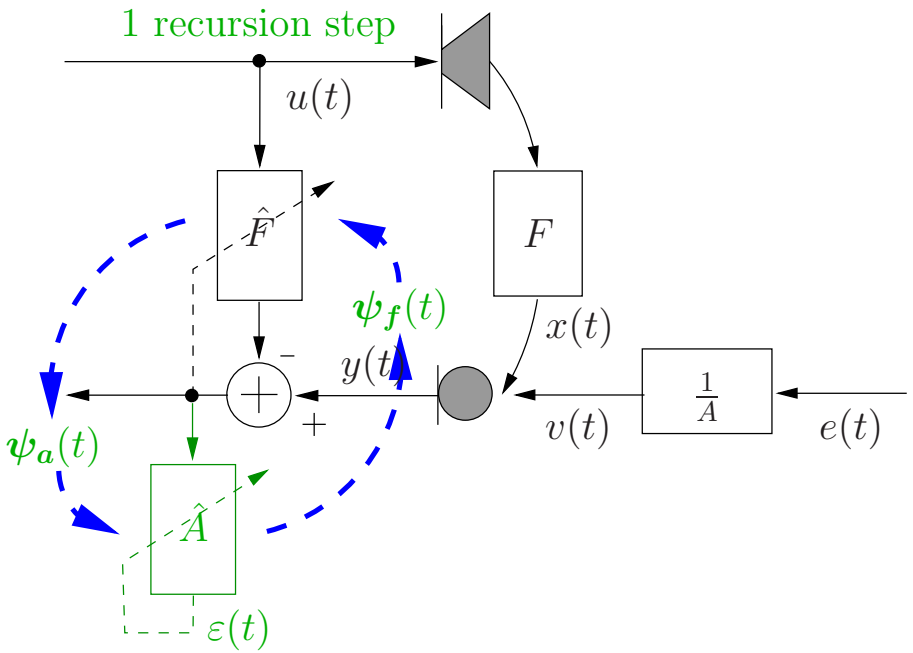
Robust AF Solution 2: PEM-AF algorithm [Spriet et al., 2004]

- Gradient vector calculation:

$$\boldsymbol{\psi_f}(t) = \begin{bmatrix} \hat{A}(q,t-1)u(t) \\ \tilde{\boldsymbol{\psi}}_{\boldsymbol{f}}(t-1) \end{bmatrix}$$

$$= \begin{bmatrix} \hat{A}(q,t-1)u(t) \\ \hat{A}(q,t-2)u(t-1) \\ \vdots \\ \hat{A}(q,t-n_F-1)u(t-n_F) \end{bmatrix}$$

$\rightarrow$ invalid approximation if $\hat{A}(q,t-1) \neq \hat{A}(q,t-2) \neq \hat{A}(q,t-n_F-1)$



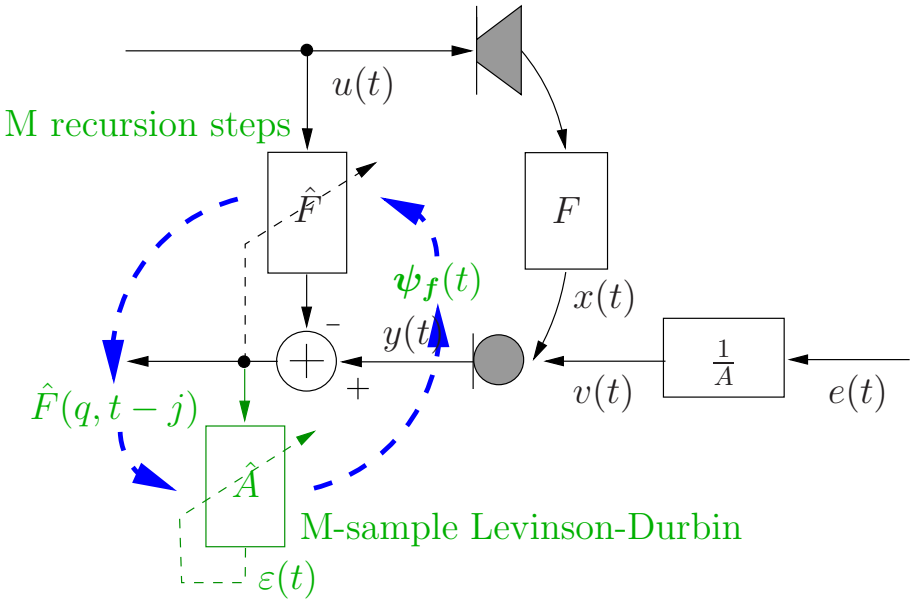
- $A(q,t)$ estimated recursively (RLS/Burg Lattice), concurrently with $F(q,t)$, using decoupled data windows

Robust AF Solution 3: PEM-AFROW algo [Rombouts et al., 2005]

- Gradient vector calculation:

$$\boldsymbol{\psi_f}(t) = \begin{bmatrix} \hat{A}(q, t-i)u(t) \\ \hat{A}(q, t-i)u(t-1) \\ \vdots \\ \hat{A}(q, t-i)u(t-n_F) \end{bmatrix}$$

$\longrightarrow$ valid approximation



- $A(q,t)$ estimated non-recursively (Levinson-Durbin) on M-sample data frames, using a previous estimate $\hat{F}(q, t-j)$

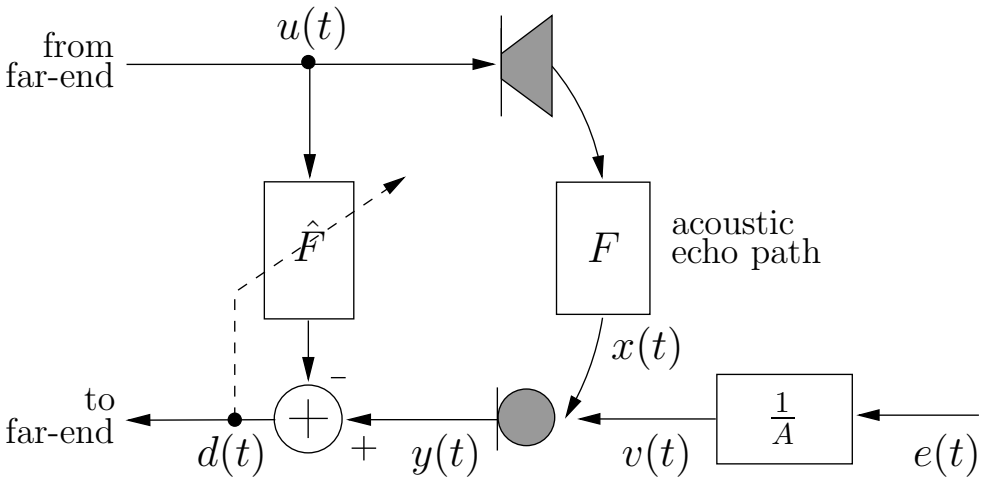
Robust AF Solution 2 & 3: PEM based adaptive filtering

- PEM-AF algorithm:
  - Pro: decoupling of the data windows for estimation of $F(q,t)$ and $A(q,t)$
  - Contra: non-convex cost function (local minima),
    approximate gradient vector calculation, assuming that $A(q,t)$ remains stationary during n_F samples

- PEM-AFROW algorithm:
  - Pro: decoupling of the data windows for estimation of $F(q,t)$ and $A(q,t)$,
    exact gradient vector calculation (only row operations)
  - Contra: non-convex cost function (local minima)

Simulation results (1)

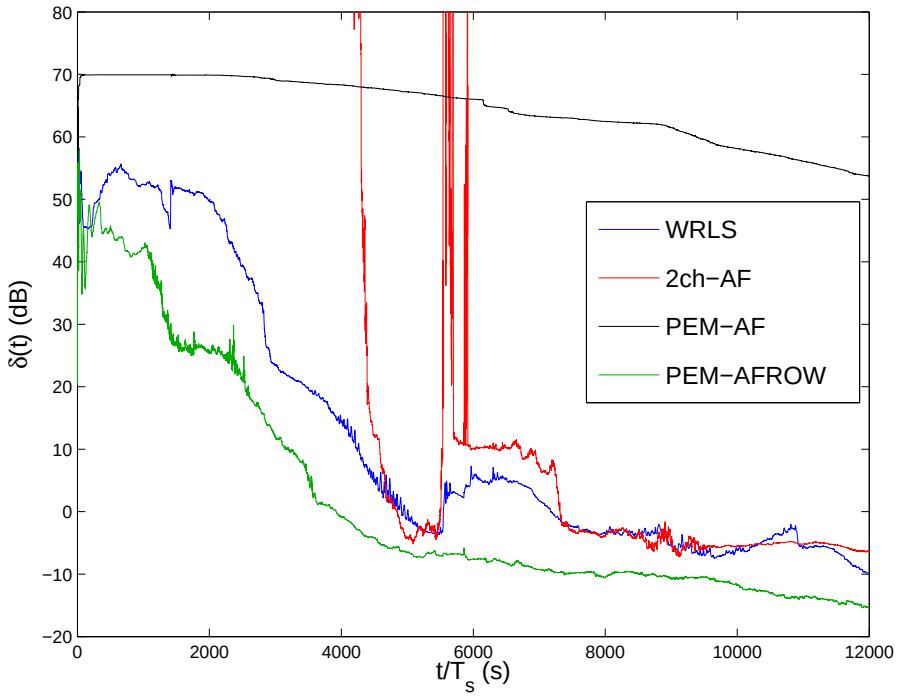
- Simulation parameters:
  - $f_s = 8kHz$
  - $n_F = 1000$
  - $n_A = 55$
  - $u(t) =$ male speech
  - $v(t) =$ female speech
  - echo-to-background ratio $\frac{\sum_{k=1}^{N} |x(k)|^2}{\sum_{k=1}^{N} |v(k)|^2} = 10dB$



- Performance measure: normalized distance $\delta(t) = 20 \log_{10} \frac{\|\hat{\mathbf{f}}(t) - \mathbf{f}\|}{\|\mathbf{f}\|}$.
- 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.

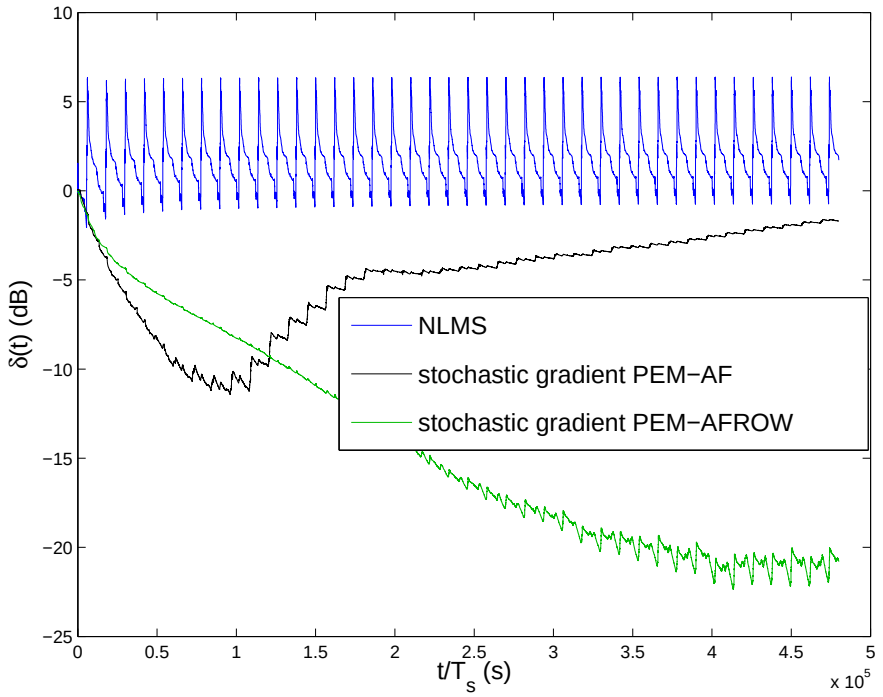
Simulation results (2)

Gauss-Newton type algorithms



RLS	4,006,006	FLOPS per recursion
2ch-AF	5,916,310	
PEM-AF	4,020,216	
PEM-AFROW	4,008,600	

Stochastic gradient algorithms



NLMS	7,001	FLOPS per recursion
2ch-AF	1,018,377	
PEM-AF	21,215	
PEM-AFROW	9,580	

Conclusions

- A continuous double-talk situation, which may occur e.g. in a noisy teleconferencing environment or in an automatic gain adjustment or feedback cancellation 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 with only little additional computational effort compared to standard RLS/NLMS.