Acoustic Echo Cancellation with Improved Double-Talk Robustness and Optimal Regularization

Toon van Waterschoot* (ESAT-SCD/SISTA)

SISTA Seminar 11/01/2005

*in collaboration with Geert Rombouts, Kris Struyve (Televic N.V.) and Marc Moonen (promotor)

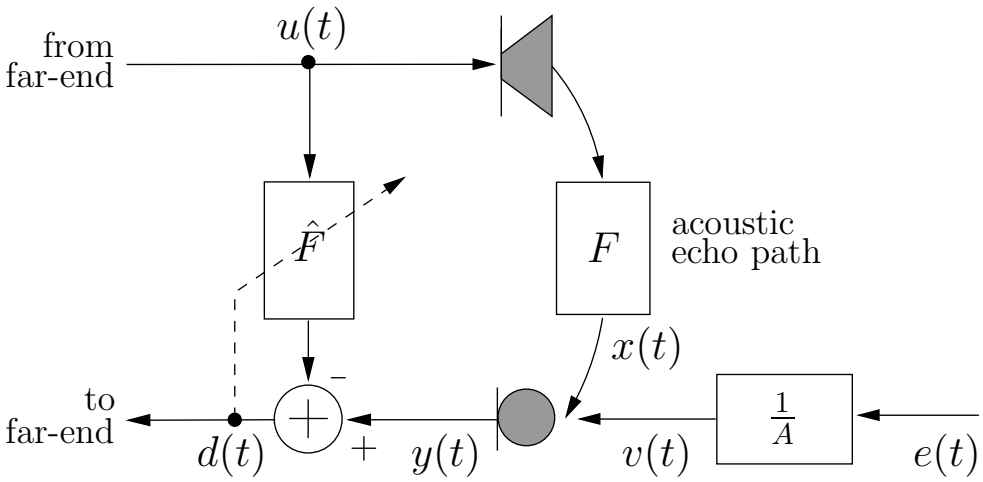
SISTA

Overview

- Acoustic Echo Cancellation
  - double-talk problem
  - poor excitation problem
- Linear Estimation
  - best linear unbiased estimator (BLUE)
  - linear minimum mean square error (LMMSE) estimator
- Recursive Identification
  - prediction error method (PEM)
  - regularized prediction error method (RgPEM)
- Simulation Results
- Conclusion

Acoustic Echo Cancellation: Concept

- loudspeaker signal from far-end $u(t)$
- near-end signal: $v(t) = \frac{1}{A(q,t)}e(t)$
- near-end excitation signal $e(t)$: white noise with variance σ_t^2
- microphone signal: $y(t) = F(q,t)u(t) + v(t)$
- echo-compensated signal to far-end: $d(t) = y(t) - \hat{F}(q,t)u(t)$



- room impulse response (RIR): $F(q,t) = f_0(t) + f_1(t)q^{-1} + \ldots + f_{n_F}(t)q^{-n_F}$
- near-end signal autoregressive (AR) model: $A(q,t) = 1 + a_1(t)q^{-1} + \ldots + a_{n_A}(t)q^{-n_A}$
- Goal: estimate and track time-varying RIR $F(q,t) \Rightarrow$ low variance estimate required

Acoustic Echo Cancellation: LS estimator

The linear estimation problem from which the RIR can be estimated, may be written as

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

$$\Updownarrow$$

$$\mathbf{y} = \mathbf{U}\mathbf{f} + \mathbf{v}.$$

Minimization of the least squares (LS) criterion

$$V_{LS}(t, \mathbf{f}) = (\mathbf{y} - \mathbf{U}\mathbf{f})^T (\mathbf{y} - \mathbf{U}\mathbf{f})$$

results in the well-known LS estimator

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

Acoustic Echo Cancellation: double-talk problem

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

- If near-end signal $v(t) \neq 0$, the LS estimator will have a large variance:

$$\begin{aligned} \text{cov}(\hat{\mathbf{f}}_{\mathbf{LS}}) &= (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{R}\mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}, \\ \mathbf{R} &\triangleq E\mathbf{v}\mathbf{v}^T, \end{aligned}$$

  or, with finite data memory, a large excess error.
- With stocastic gradient (e.g. NLMS) and affine projection adaptive algorithms, double-talk may even lead to divergence.
- Standard solution is to include a double-talk detector that freezes the adaptation during double-talk.
- Still:
  - additional robustness is required to avoid divergence during detection interval,
  - no adequate solution for 'continuous double-talk situation' (e.g. when near-end signal is background noise).

Acoustic Echo Cancellation: poor excitation problem

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

- Matrix $\mathbf{U}^T\mathbf{U}$ may be very badly conditioned, because:
  - audio signals may exhibit a nearly harmonic spectrum,
  - $\dim(\mathbf{U}^T\mathbf{U}) = (n_F + 1) \times (n_F + 1)$ with e.g. $n_F = 1000$ at $f_s = 8kHz$.
- Standard solution: Tikhonov regularization (related to ridge regression),

$$\hat{\mathbf{f}}_{\mathbf{RgLS}} = (\mathbf{U}^T\mathbf{U} + \alpha\mathbf{I})^{-1}\mathbf{U}^T\mathbf{y}.$$

- Still:
  - rather ad-hoc solution, probably suboptimal,
  - how to choose regularization parameter α?

Linear Estimation: BLUE

- Any linear estimator of $\mathbf{f}$ can be written as a linear function of the microphone signal vector $\mathbf{y}$:

$$\hat{\mathbf{f}} = \mathbf{Z}^T \mathbf{y}.$$

- Minimizing the variance $E(\hat{\mathbf{f}} - E\hat{\mathbf{f}})(\hat{\mathbf{f}} - E\hat{\mathbf{f}})^T$ under the unbiasedness constraint $\mathbf{Z}^T\mathbf{U} = \mathbf{I}$ results in the best linear unbiased estimator (BLUE):

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

- Indeed, taking into account the near-end signal characteristics ($\mathbf{R} = E\mathbf{v}\mathbf{v}^T$) will lower the variance of the RIR estimate.

Linear estimation: LMMSE estimator (1)

- The variance can be lowered even further by omitting the unbiasedness constraint.
- A criterion in which the variance and the squared bias are weighted equally is the minimum mean square error (MMSE) criterion:

$$V_{MMSE}(t, \hat{\mathbf{f}}) = E(\hat{\mathbf{f}} - \mathbf{f_0})(\hat{\mathbf{f}} - \mathbf{f_0})^T.$$

  where $\mathbf{f_0}$ represents the true RIR.
- Bayesian framework: think of $\mathbf{f_0}$ as a random variable on which some prior knowledge may be available, i.e. suppose $\mathbf{f_0}$ has a prior PDF $p(\mathbf{f_0})$ with first and second order moments:

$$\begin{aligned} \boldsymbol{\mu_0} &\triangleq E\mathbf{f_0} \\ \mathbf{P_0} &\triangleq \text{cov}(\mathbf{f_0}) = E(\mathbf{f_0} - E\mathbf{f_0})(\mathbf{f_0} - E\mathbf{f_0})^T. \end{aligned}$$

Linear estimation: LMMSE estimator (2)

- The linear estimator $\hat{\mathbf{f}} = \mathbf{Z}^T\mathbf{y}$ minimizing $V_{MMSE}(t, \hat{\mathbf{f}})$ is then given by the mean of the posterior PDF $p(\mathbf{f_0}|\mathbf{y})$:

$$\hat{\mathbf{f}}_{\mathbf{LMMSE}} = E(\mathbf{f_0}|\mathbf{y}) = \boldsymbol{\mu_0} + (\mathbf{U}^T\mathbf{R}^{-1}\mathbf{U} + \mathbf{P_0}^{-1})^{-1}\mathbf{U}^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{U}\boldsymbol{\mu_0}).$$

  Or, if prior knowlegde is constructed such that $\boldsymbol{\mu_0} = \mathbf{0}$,

$$\boxed{\hat{\mathbf{f}}_{\mathbf{LMMSE}} = (\mathbf{U}^T\mathbf{R}^{-1}\mathbf{U} + \mathbf{P_0}^{-1})^{-1}\mathbf{U}^T\mathbf{R}^{-1}\mathbf{y}.}$$

- Compare with:

$$\begin{aligned} \hat{\mathbf{f}}_{\mathbf{BLUE}} &= (\mathbf{U}^T\mathbf{R}^{-1}\mathbf{U})^{-1}\mathbf{U}^T\mathbf{R}^{-1}\mathbf{y} \\ \hat{\mathbf{f}}_{\mathbf{RgLS}} &= (\mathbf{U}^T\mathbf{U} + \alpha\mathbf{I})^{-1}\mathbf{U}^T\mathbf{y} \end{aligned}$$

Recursive identification: 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.

Recursive identification: PEM-AFROW algorithm

One possibility to minimize $V_{PE}(t, A(q,t), F(q,t))$ recursively, is to apply the PEM-AFROW algorithm developed for adaptive feedback cancellation (see previous talk):

1. estimate $A(q,t)$ and σ_t^2 on frames with length $20ms$,
2. prefilter loudspeaker and microphone signals with $A(q,t)$:

$$\begin{aligned} u_A(t) &= A(q,t)u(t), \\ y_A(t) &= A(q,t)y(t), \end{aligned}$$

3. estimate RIR $\mathbf{f}(t)$ recursively:

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

Recursive identification: regularized PEM

- The LMMSE estimator may be obtained by minimizing the regularized prediction error criterion

$$V_{RgPE}(t) = \frac{1}{2N}\left(\sum_{k=1}^{t}\frac{\lambda^{t-k}}{\hat{\sigma}_k^2}\big(A(q,t)[y(k) - F(q,t)u(k)]\big)^2 + \mathbf{f}^T(t){\mathbf{P_0}}^{-1}\mathbf{f}(t)\right).$$

- The recursive update equations for $\mathbf{f}(t)$ in the PEM-AFROW algorithm then look like

$$\begin{aligned}
\hat{\mathbf{f}}(t) &= \big(\mathbf{I} - (1-\lambda)\mathbf{R}^{-1}(t){\mathbf{P_0}}^{-1}\big)\hat{\mathbf{f}}(t-1) + \frac{1}{\hat{\sigma}_t^2}\mathbf{R}^{-1}(t)\mathbf{u}_A(t)\varepsilon(t),\\
\mathbf{R}(t) &= \lambda\mathbf{R}(t-1) + \frac{1}{\hat{\sigma}_t^2}\mathbf{u}_A(t)\mathbf{u}_A^T(t) + (1-\lambda){\mathbf{P_0}}^{-1},\\
\varepsilon(t) &= y_A(t) - \mathbf{u}_A^T(t)\hat{\mathbf{f}}(t-1).
\end{aligned}$$

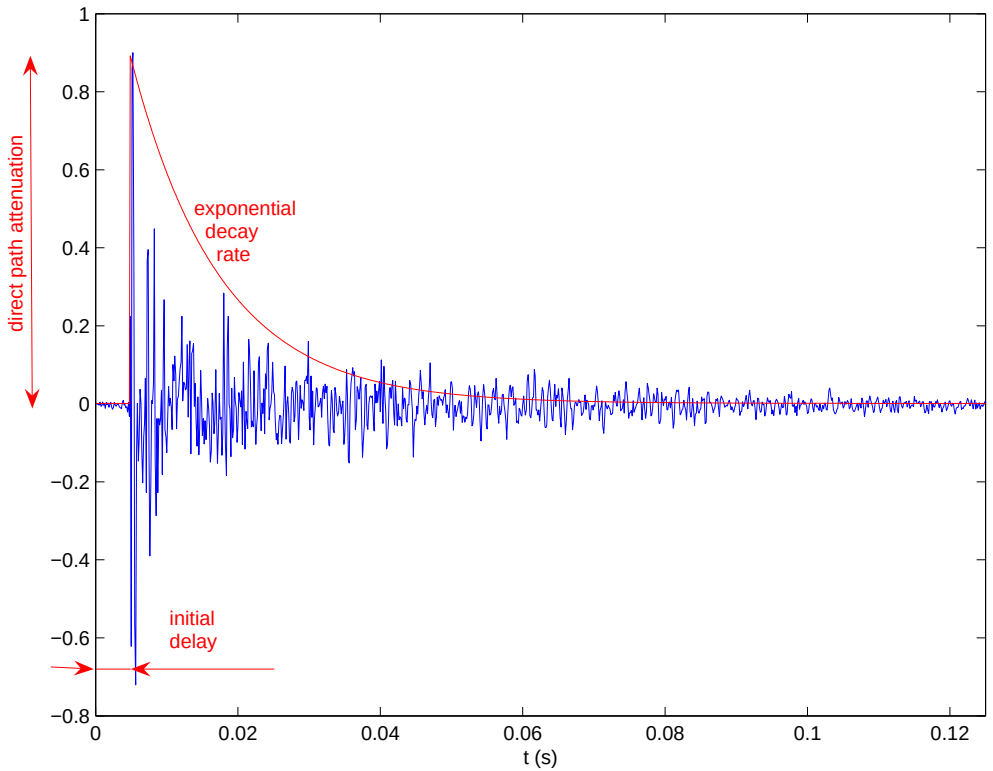
- Note:
  - with ${\mathbf{P_0}}^{-1} = \alpha I$ we get the ridge regression criterion,
  - the matrix inversion lemma cannot be applied to the above update equation for $\mathbf{R}(t)$.

Recursive identification: choice of $\mathbf{P_0}^{-1}$

- Construct $\mathbf{P_0}^{-1}$ from a priori RIR estimate $\bar{\mathbf{f}} = [\bar{f}_0 \;\; \bar{f}_1 \;\; \dots \;\; \bar{f}_{n_F}]^T$, e.g. as a diagonal matrix

$$\mathbf{P_0}^{-1} \triangleq \begin{bmatrix} \frac{1}{\bar{f}_0^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\bar{f}_1^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\bar{f}_{n_F}^2} \end{bmatrix}.$$

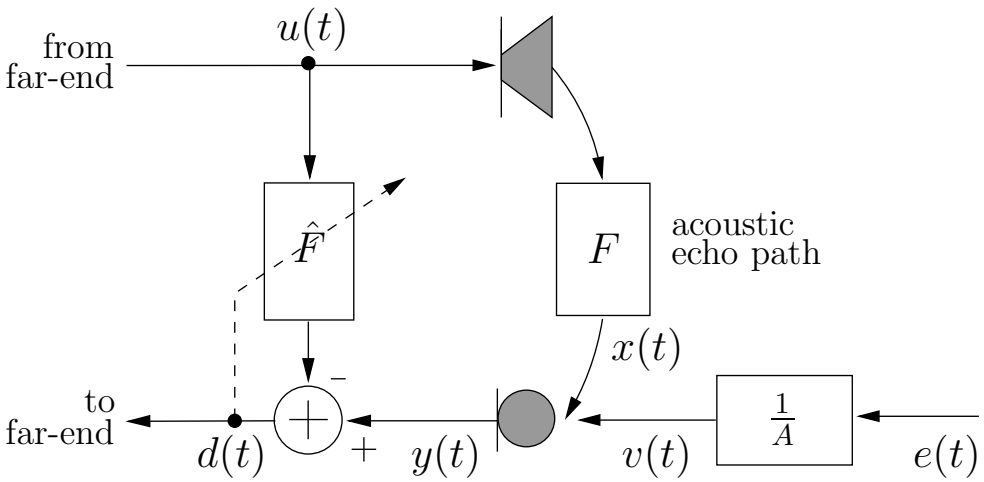
The a priori estimate $\bar{\mathbf{f}}$ may be measured during system setup or calculated from physical parameters (e.g. initial delay, direct path attenuation and exponential decay rate).



- Estimate $\mathbf{P_0}^{-1}$ from previous RIR estimates: $\mathbf{P_0}^{-1}(t) = \lambda \mathbf{P_0}^{-1}(t-1) + (1-\lambda)\mathbf{f}(t)\mathbf{f}^T(t)$.

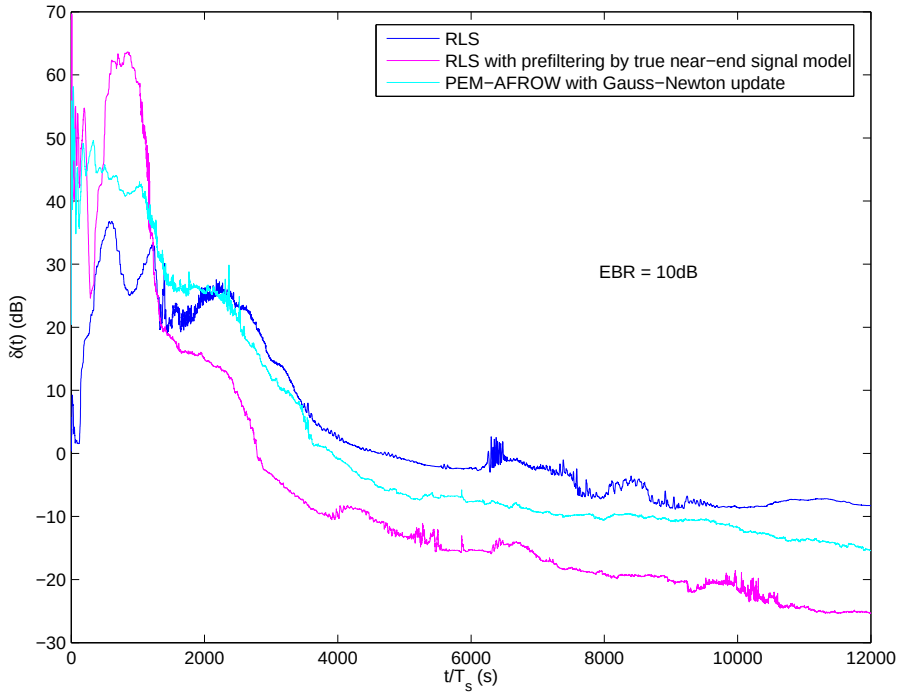
Simulation results (1)

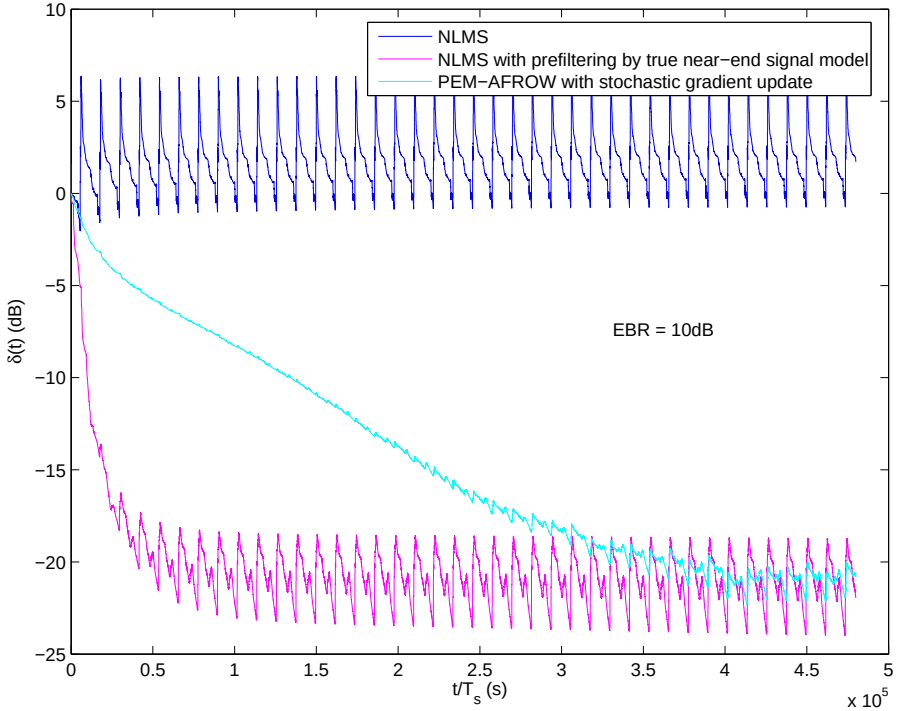
- Simulation parameters:
  - $f_s = 8kHz$
  - $n_F = 1000$
  - $n_A = 55$
  - $u(t) =$ male speech
  - $v(t) =$ female speech
  - echo-to-background ratio $\frac{\sum_{k=1}^{N} |x(k)|^2}{\sum_{k=1}^{N} |v(k)|^2} = 10dB$



- Performance measure: $\delta(t) = 20 \log_{10} \frac{\|\hat{\mathbf{f}}(t) - \mathbf{f}\|}{\|\mathbf{f}\|}$.
- Three algorithms have been compared in two flavours:
  - with Gauss-Newton update: RLS, RLS with $A(q,t)$ and σ_t^2 estimated on clean near-end signal, PEM-AFROW,
  - with stochastic gradient update: NLMS, NLMS with $A(q,t)$ and σ_t^2 estimated on clean near-end signal, PEM-AFROW.

Simulation results (2)





Conclusion

- Some nice results from linear estimation theory may be applied to the problem of double-talk (BLUE) and poor excitation (LMMSE estimator) occuring in acoustic echo cancellation.
- Using recursive prediction error identification theory these estimators may be implemented in an adaptive algorithm, e.g. the PEM-AFROW algorithm.
- The complexity increase due to the estimation of the near-end signal model is reasonable (note that in the simulations $n_A = 55$ compared to $n_F = 1000$). In the (optimally) regularized adaptive algorithms however, complexity will rise because the matrix inversion lemma cannot be applied.
- From simulation results we conclude that double-talk robustness may be increased by prefiltering with the inverse near-end signal model and weighting with the inverse near-end excitation signal variance. However, convergence to a local minimum of the prediction error criterion may decrease the performance of the proposed method.