Practical Variable Step-Size Adaptive Algorithms for Echo Cancellation

Constantin Paleologu[1], *Jacob Benesty*[2], *Silviu Ciochină*[1], *and Călin Vlădeanu*[1]

[1]Telecommunications Department
University Politehnica of Bucharest, Romania
{pale, silviu, calin}@comm.pub.ro
[2]INRS-EMT
Universite du Quebec, Montreal, Canada
benesty@emt.inrs.ca

Introduction

- **adaptive echo cancellation**

→ an *adaptive filter* identifies the echo path

- resulted due to the hybrid circuit (2 wire ←→ 4 wire)

[*network echo cancellation* (NEC)]

- between the terminal's loudspeaker and microphone

[*acoustic echo cancellation* (AEC)]

- **specific problems in echo cancellation**

→ the echo path can be extremely long

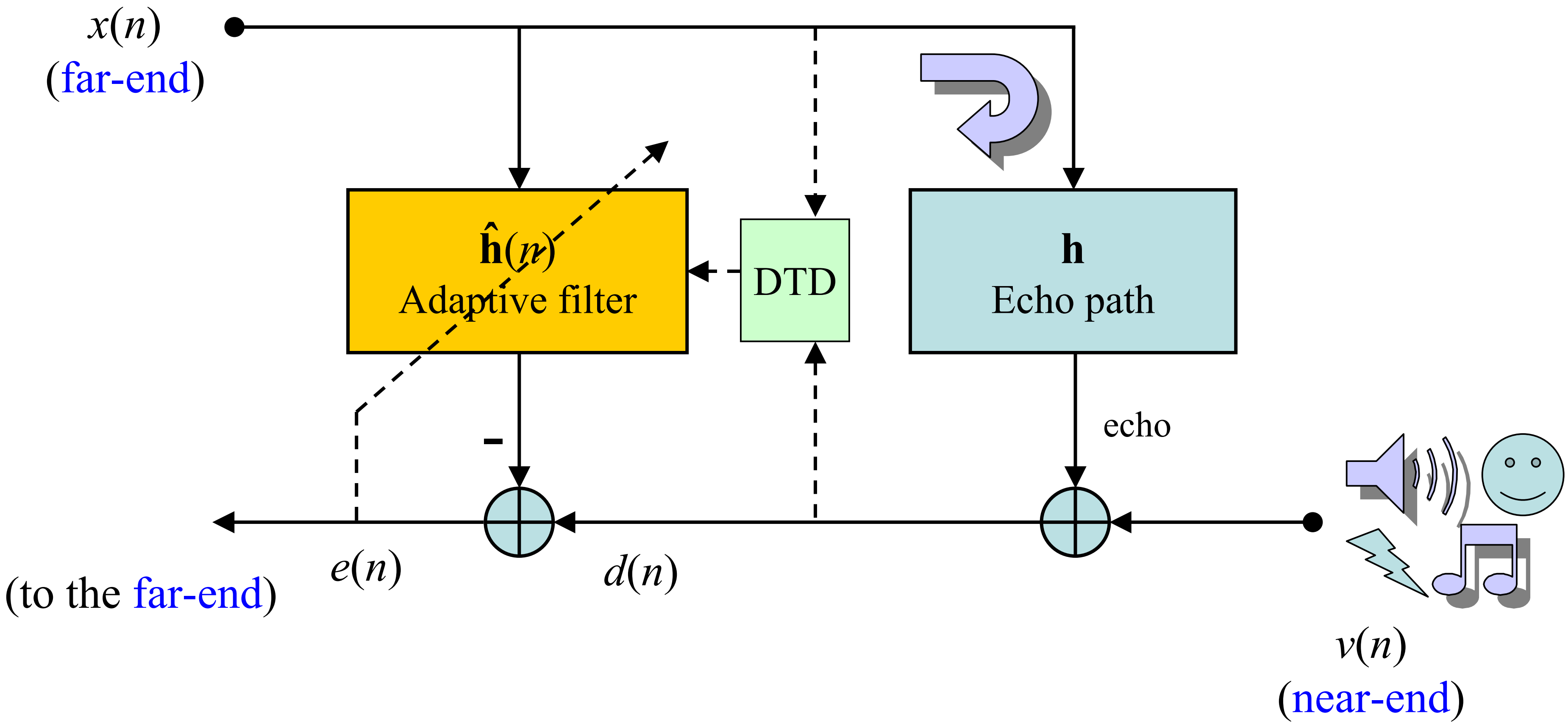
→ it may rapidly change at any time during the connection

→ the background noise can be strong and non-stationary

→ the behaviour during double-talk

→ the presence of the Double-Talk Detector (DTD)

Basic configuration



Adaptive algorithms for echo cancellation

- **requirements**

  → fast convergence rate and tracking

  → low misadjustment

  → double-talk robustness

- **common choices**

  → Normalized least-mean-square (NLMS) algorithm

  → Affine Projection Algorithm (APA)

- **step-size parameter**

  → *large* values **→** fast convergence rate and tracking

  → *small* values **→** low misadjustment and double-talk robustness

conflicting requirements **→ variable step-size (VSS) algorithms**

- **classical NLMS algorithm**

$$\boxed{\begin{aligned} e(n) &= d(n) - \mathbf{x}^T(n)\hat{\mathbf{h}}(n-1) \\ \hat{\mathbf{h}}(n) &= \hat{\mathbf{h}}(n-1) + \mu(n)\mathbf{x}(n)e(n) \end{aligned}}$$

step-size parameter

adaptive filter length

where $\mu(n) = \mu \,/\, [\mathbf{x}^T(n)\mathbf{x}(n)]$

$$\mathbf{x}(n) = [x(n), x(n-1), \ldots, x(n-L+1)]^T$$

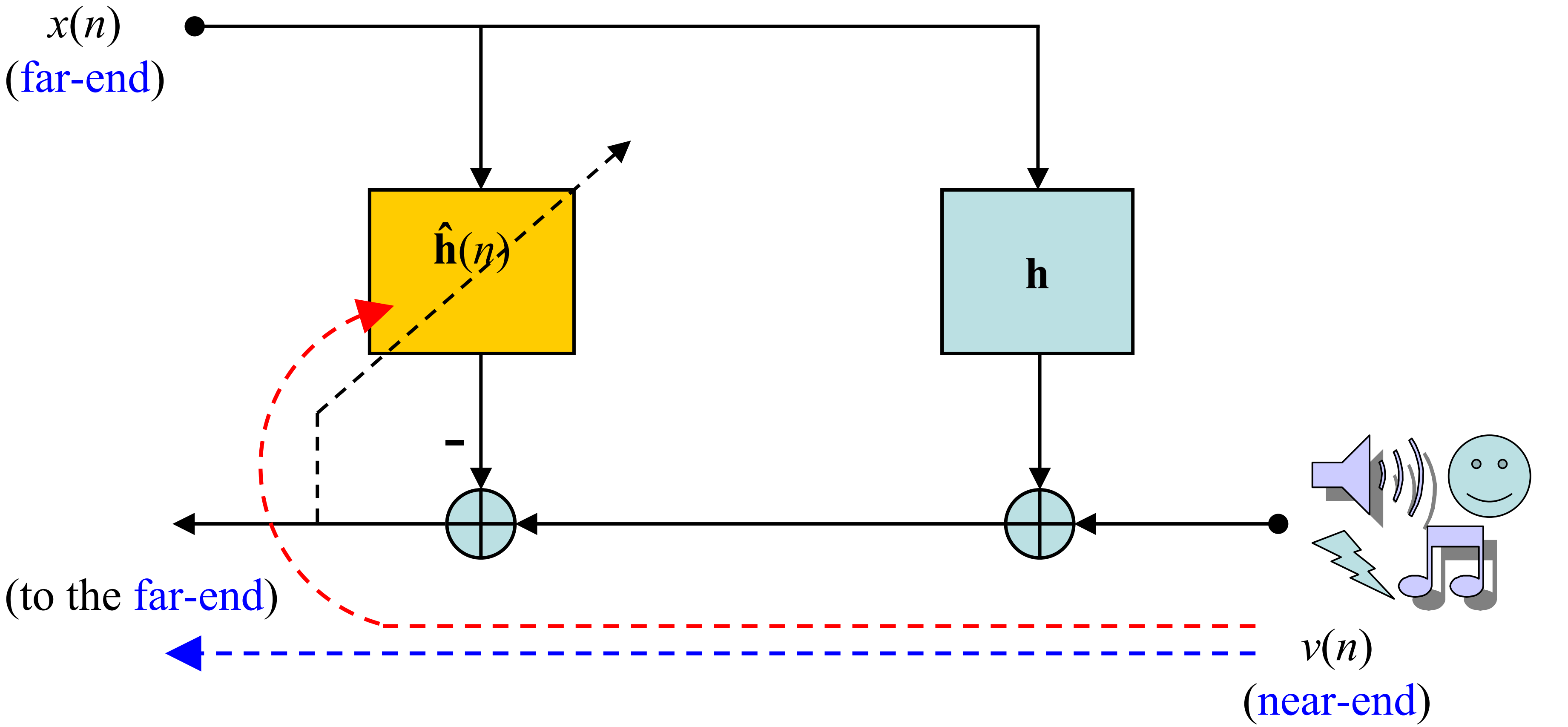
$\varepsilon(n) = d(n) - \mathbf{x}^T(n)\hat{\mathbf{h}}(n)$ → *a posteriori* error vector

➔ $\varepsilon(n) = e(n)[1 - \mu(n)\mathbf{x}^T(n)\mathbf{x}(n)]$

$\varepsilon(n) = 0$ [assuming that $e(n) \neq 0$] ➔ $\mu(n) = 1 \,/\, \mathbf{x}^T(n)\mathbf{x}(n)]$

➔ $\mu = 1$

Optimal step-size ?



$\cancel{\varepsilon(n) = 0}$

$\boxed{\varepsilon(n) = v(n)}$

$$E\{\varepsilon^2(n)\} = E\{v^2(n)\}$$

$$E\{e^2(n)\}[1 - \mu(n) L E\{x^2(n)\}]^2 = E\{v^2(n)\}$$

$$\Rightarrow \quad \mu(n) = \frac{1}{\mathbf{x}^T(n)\mathbf{x}(n)}\left[1 - \sqrt{\frac{E\{v^2(n)\}}{E\{e^2(n)\}}}\right]$$

$$\boxed{\mu(n) = \frac{1}{\mathbf{x}^T(n)\mathbf{x}(n)}\left[1 - \frac{\hat{\sigma}_v(n)}{\hat{\sigma}_e(n)}\right]}$$

?

$$\hat{\sigma}_\alpha^2(n) = \lambda \hat{\sigma}_\alpha^2(n-1) + (1-\lambda)\alpha^2(n)$$

$\lambda = 1 - 1/(KL)$, with $K > 1$

1) near-end signal = background noise (*single-talk scenario*)

$$v(n) = w(n)$$

background noise power estimate

$$\mu(n) = \frac{1}{\mathbf{x}^T(n)\mathbf{x}(n)}\left[1 - \frac{\hat{\sigma}_w}{\hat{\sigma}_e(n)}\right]$$

[J. Benesty *et al*, "A nonparametric VSS NLMS algorithm", *IEEE Signal Process. Lett.*, 2006]

Problem: background noise can be time-variant

2) near-end signal = background noise + near-end speech

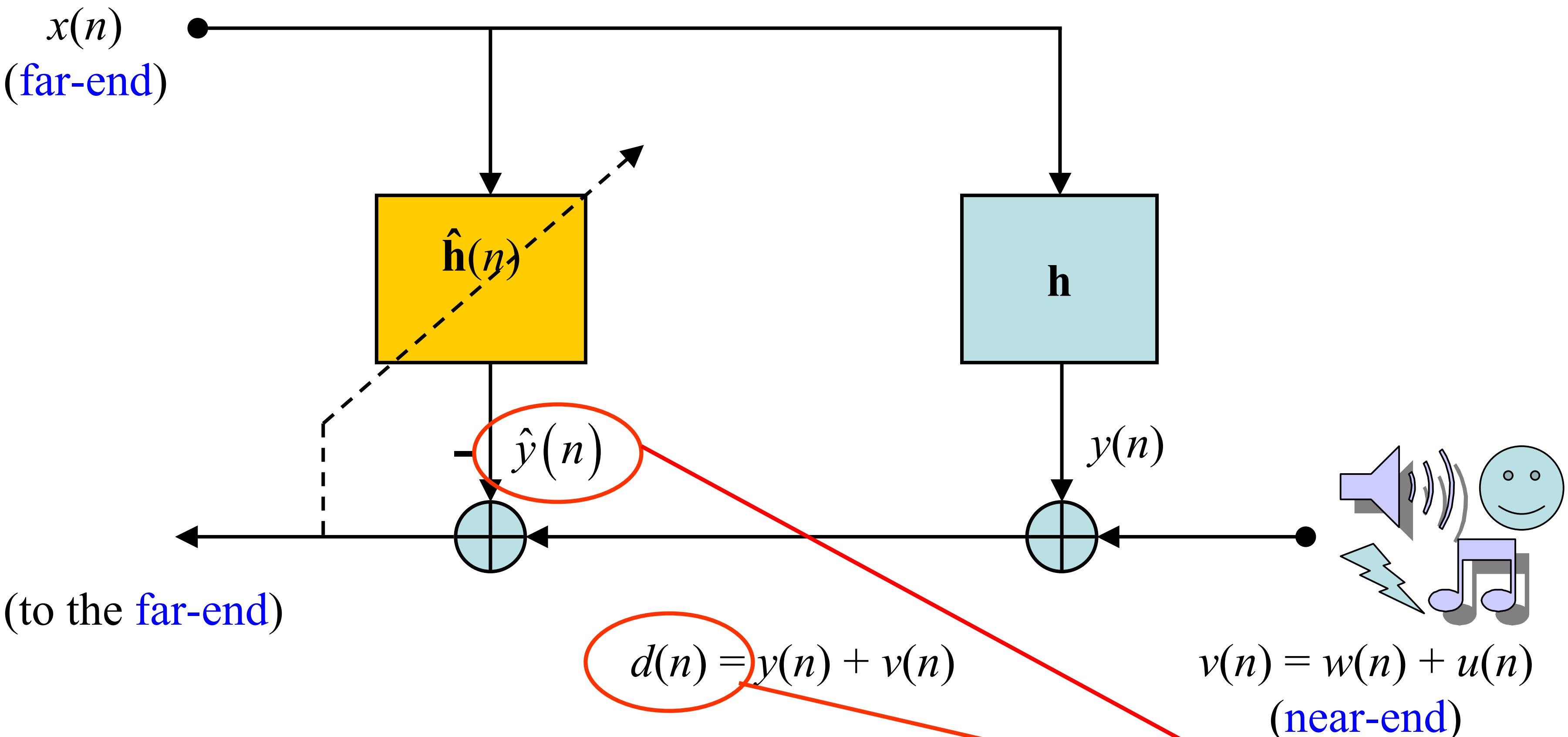
$$v(n) = w(n) + u(n) \qquad (\textit{double-talk scenario})$$

near-end speech power estimate

$$\hat{\sigma}_v^2(n) = \hat{\sigma}_w^2(n) + \hat{\sigma}_u^2(n)$$

???

Problem: non-stationary character of the speech signal



available !

$$\begin{cases} E\{d^2(n)\} = E\{y^2(n)\} + E\{v^2(n)\} \\ E\{y^2(n)\} \cong E\{\hat{y}^2(n)\} \end{cases}$$

→ assuming that the adaptive filter has converged to a certain degree

$$E\{v^2(n)\} \cong E\{d^2(n)\} - E\{\hat{y}^2(n)\} \quad \Rightarrow \quad \boxed{\hat{\sigma}_v^2(n) \cong \hat{\sigma}_d^2(n) - \hat{\sigma}_{\hat{y}}^2(n)}$$

$$\boxed{\mu(n)=\frac{1}{\mathbf{x}^T(n)\mathbf{x}(n)}\left[1-\frac{\sqrt{\hat{\sigma}_d^2(n)-\hat{\sigma}_{\hat{y}}^2(n)}}{\hat{\sigma}_e(n)}\right]}$$

Proposed VSS-NLMS

- **main advantages**

  → non-parametric algorithm

  → robustness to background noise variations and double-talk

- **the idea can be extended to the APA case** ➔ **VSS-APA**

  → *superior convergence rate* as compared to the VSS-NLMS algorithm

  → *lower complexity* as compared to the recursive least-squares (RLS) algorithm

- **classical APA**

$$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{X}^T(n)\hat{\mathbf{h}}(n-1)$$

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mu\mathbf{X}(n)\left[\mathbf{X}^T(n)\mathbf{X}(n)\right]^{-1}\mathbf{e}(n)$$

where $\mathbf{d}(n) = [d(n), d(n-1), \ldots, d(n-p+1)]^T$ projection order

$\mathbf{X}(n) = [\mathbf{x}(n), \mathbf{x}(n-1), \ldots, \mathbf{x}(n-p+1)]$

$\mathbf{x}(n-l) = [x(n-l), x(n-l-1), \ldots, x(n-l-L+1)]^T$

$l = 0,1,\ldots,p-1$

adaptive filter length

Proposed VSS-APA

$$\mu_l(n) = 1 - \frac{\sqrt{\hat{\sigma}_d^2(n-l) - \hat{\sigma}_{\hat{y}}^2(n-l)}}{\hat{\sigma}_{e_{l+1}}(n)}$$

$$\boldsymbol{\mu}(n) = \mathrm{diag}\left\{\mu_0(n), \mu_1(n), \ldots, \mu_{p-1}(n)\right\}$$

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mathbf{X}(n)\left[\mathbf{X}^T(n)\mathbf{X}(n)\right]^{-1}\boldsymbol{\mu}(n)\mathbf{e}(n)$$

Simulation results

- **conditions**

  → AEC context, $L = 512$

  → $x(n)$, $u(n)$ = speech sequences

  → $w(n)$ = independent white Gaussian noise signal (SNR = 20dB)

  → normalized misalignment (dB) = $20\log_{10}(\| \mathbf{h} - \hat{\mathbf{h}}(n) \| / \| \mathbf{h} \|)$

- **algorithms for comparisons**

  → classical APA, $\mu = 0.2$

  → variable regularized APA (VR-APA)

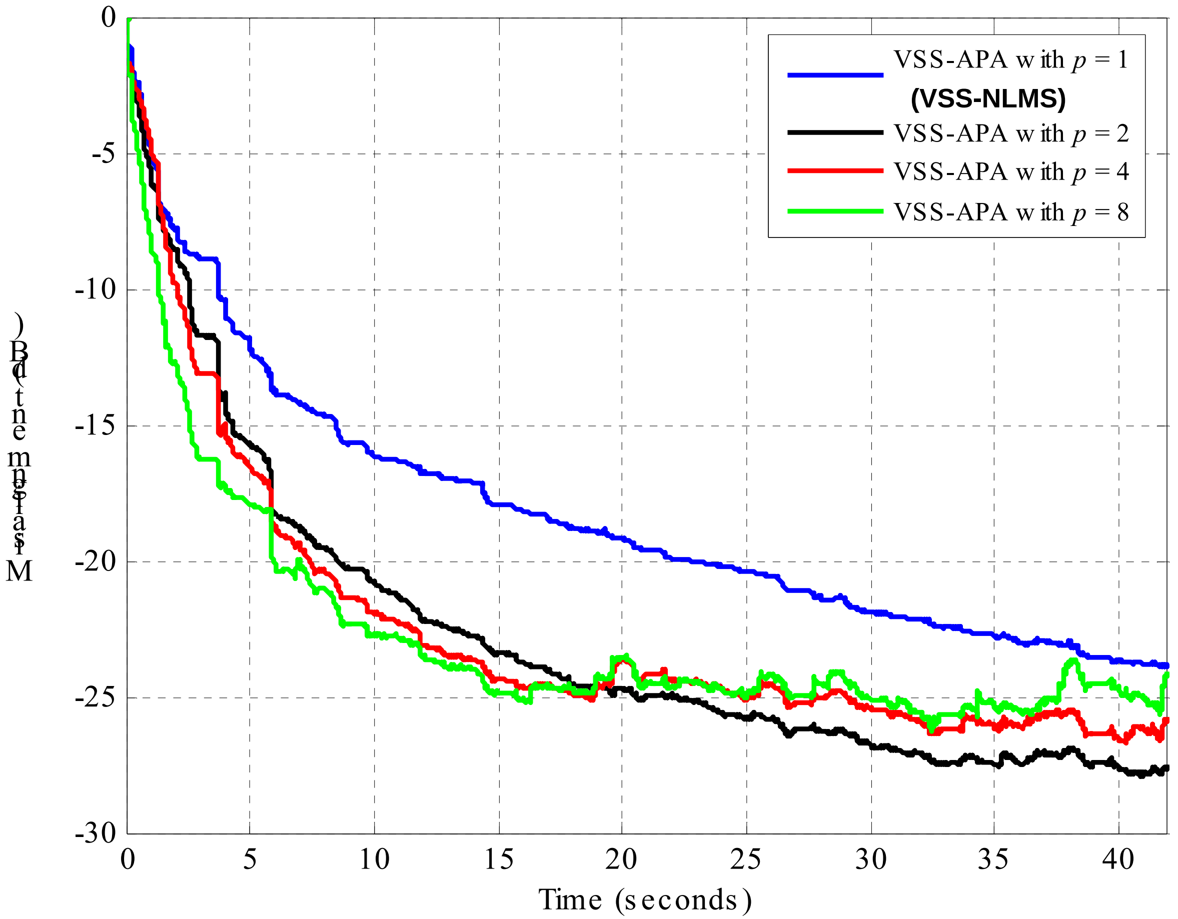
  [H. Rey, L. Rey Vega, S. Tressens, and J. Benesty, *IEEE Trans. Signal Process.*, May 2007]

  → robust proportionate APA (R-PAPA)

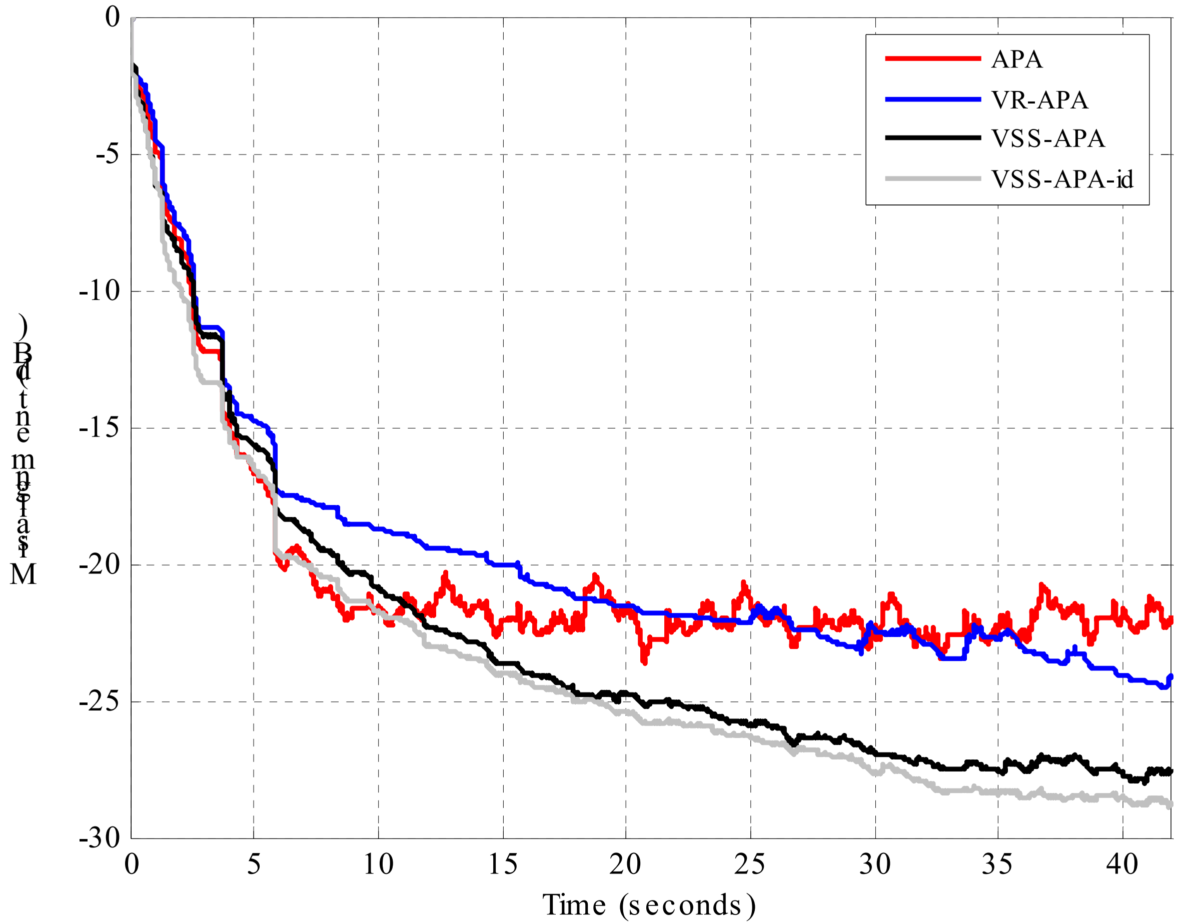
  [T. Gänsler, S. L. Gay, M. M. Sondhi, and J. Benesty, *IEEE Trans. Speech Audio Process.*, Nov. 2000]

  → "ideal" VSS-APA (VSS-APA-id) - assuming that $v(n)$ is available

- single-talk scenario (near-end signal = background noise)

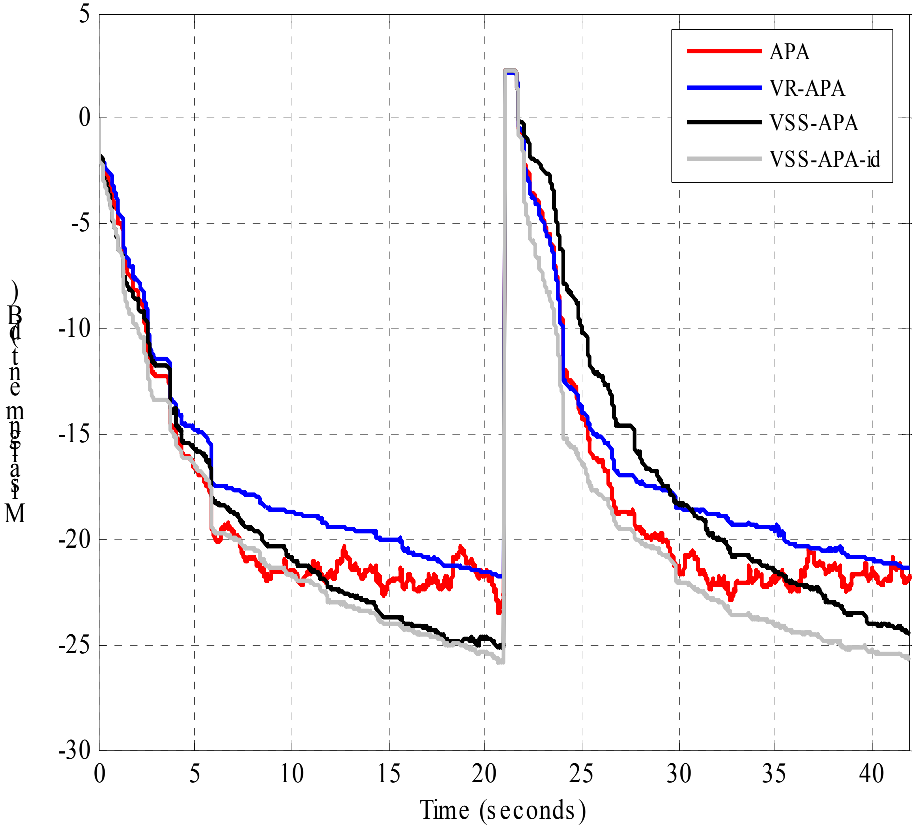


- single-talk scenario (near-end signal = background noise), $p = 2$

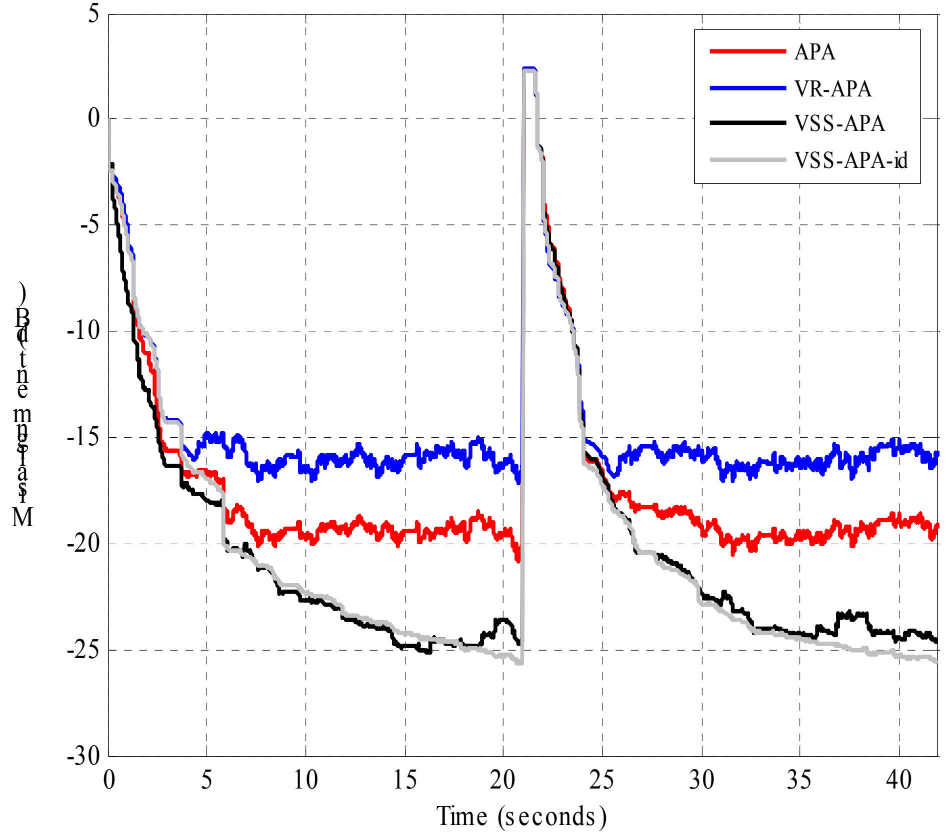


- single-talk scenario + echo path changes

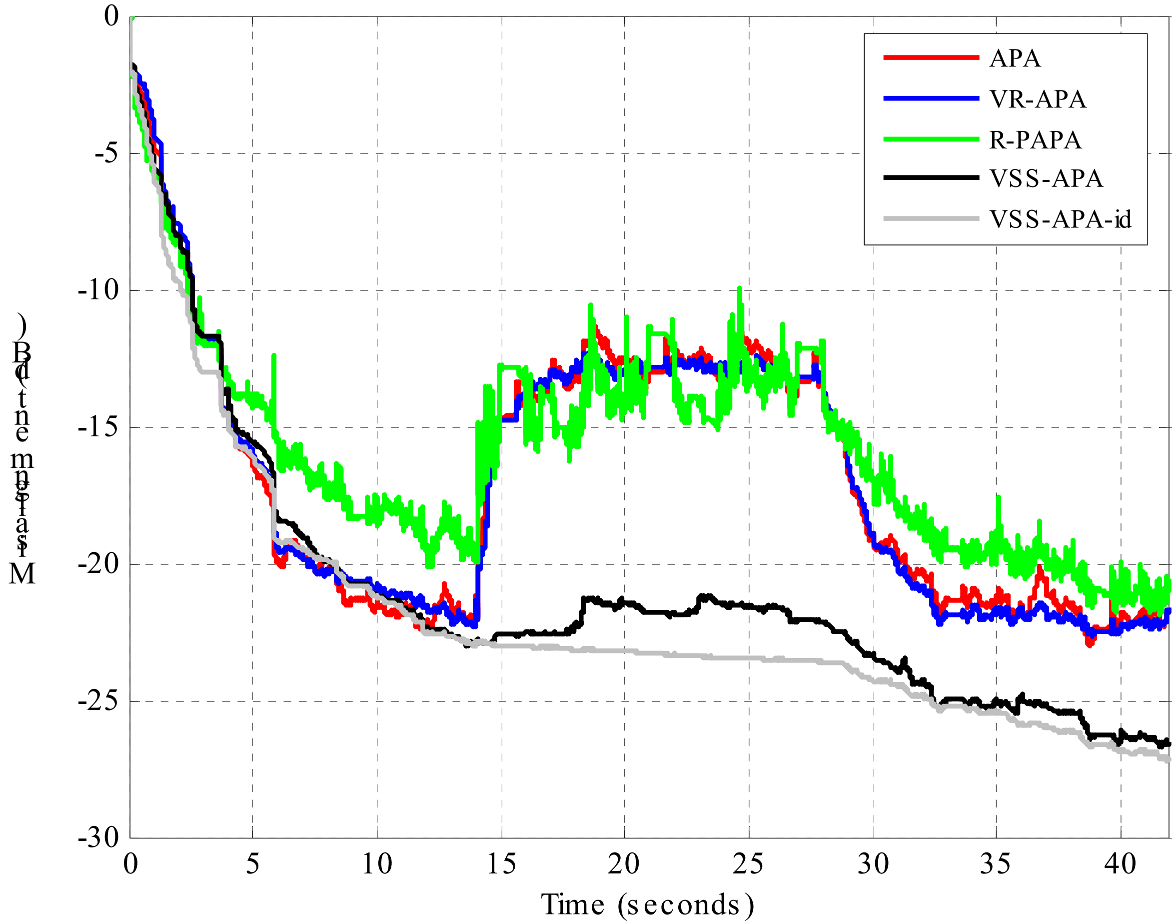
$p = 2$



$p = 8$

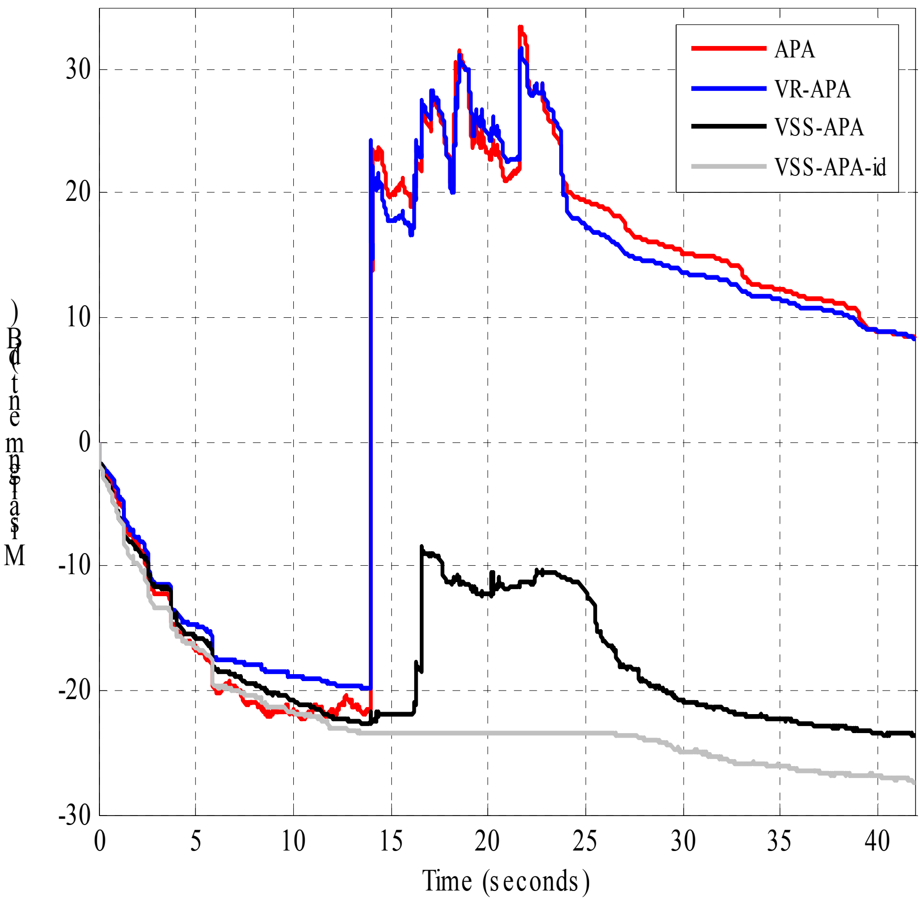


- single-talk scenario + background-noise increases (SNR → 20dB ↓ 10dB)

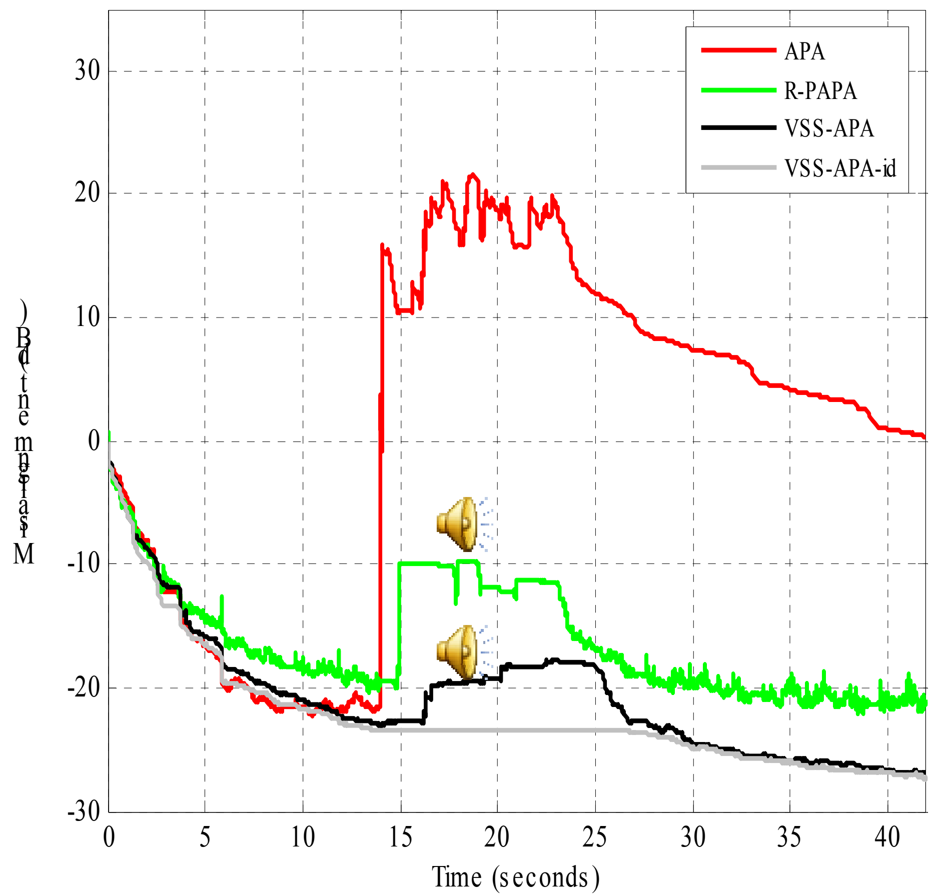


- double-talk scenario (near-end signal = background noise + near-end speech)

without DTD



with Geigel DTD



Conclusions

- VSS-NLMS and VSS-APA were developed in the context of echo cancellation.
- they takes into account the existence of the near-end signal
- the VSS formula does not require any additional parameters from the acoustic environment (i.e., non-parametric)
- they are robust to near-end signal variations like the increase of the background noise or double-talk