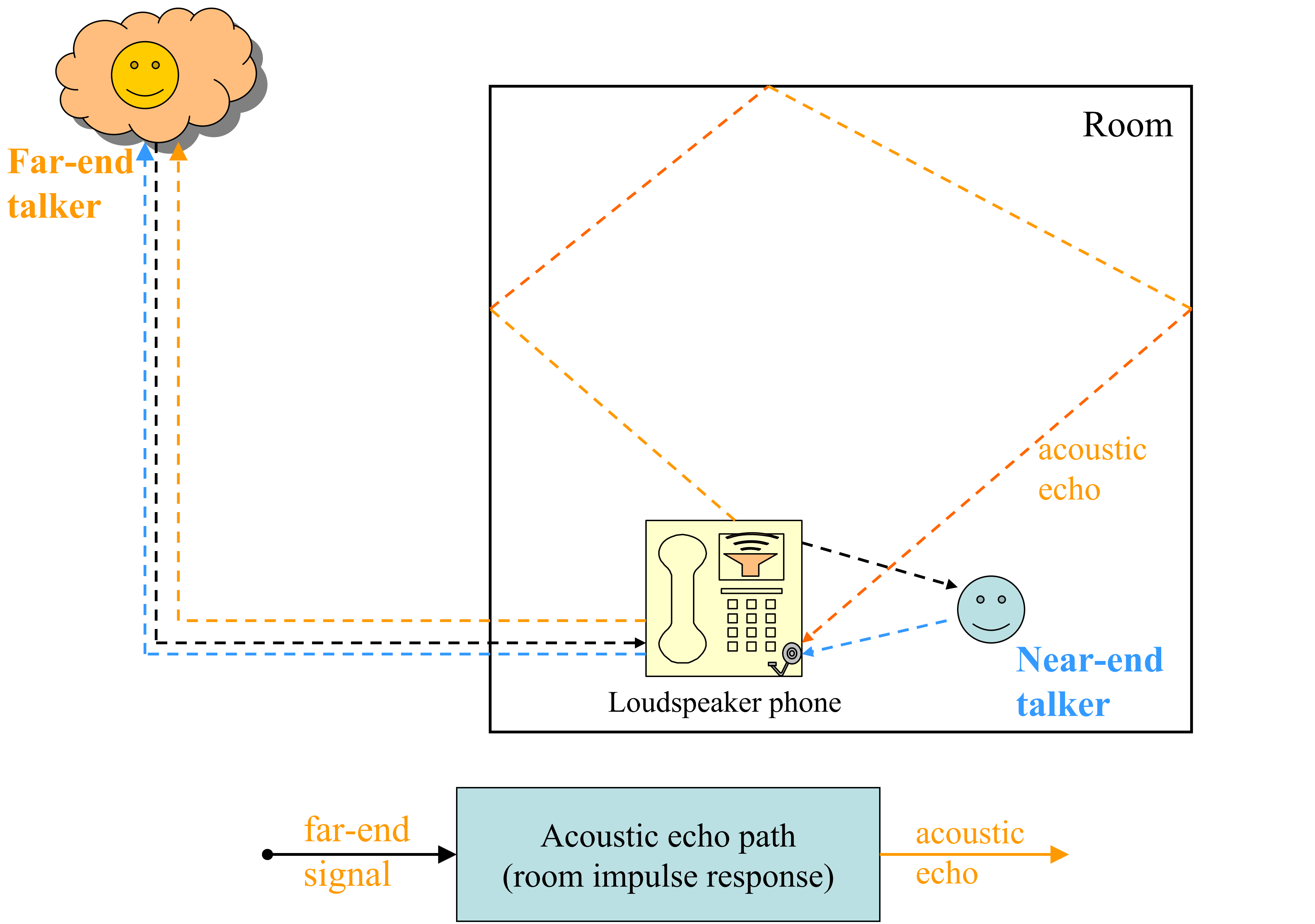
Acoustic Echo Cancellation. Challenges and Perspectives

Constantin Paleologu

Department of Telecommunications
University Politehnica of Bucharest, Romania
pale@comm.pub.ro


Far-end
talker
Room
acoustic
echo
Loudspeaker phone
Near-end
talker
far-end
signal
Acoustic echo path
(room impulse response)
acoustic
echo

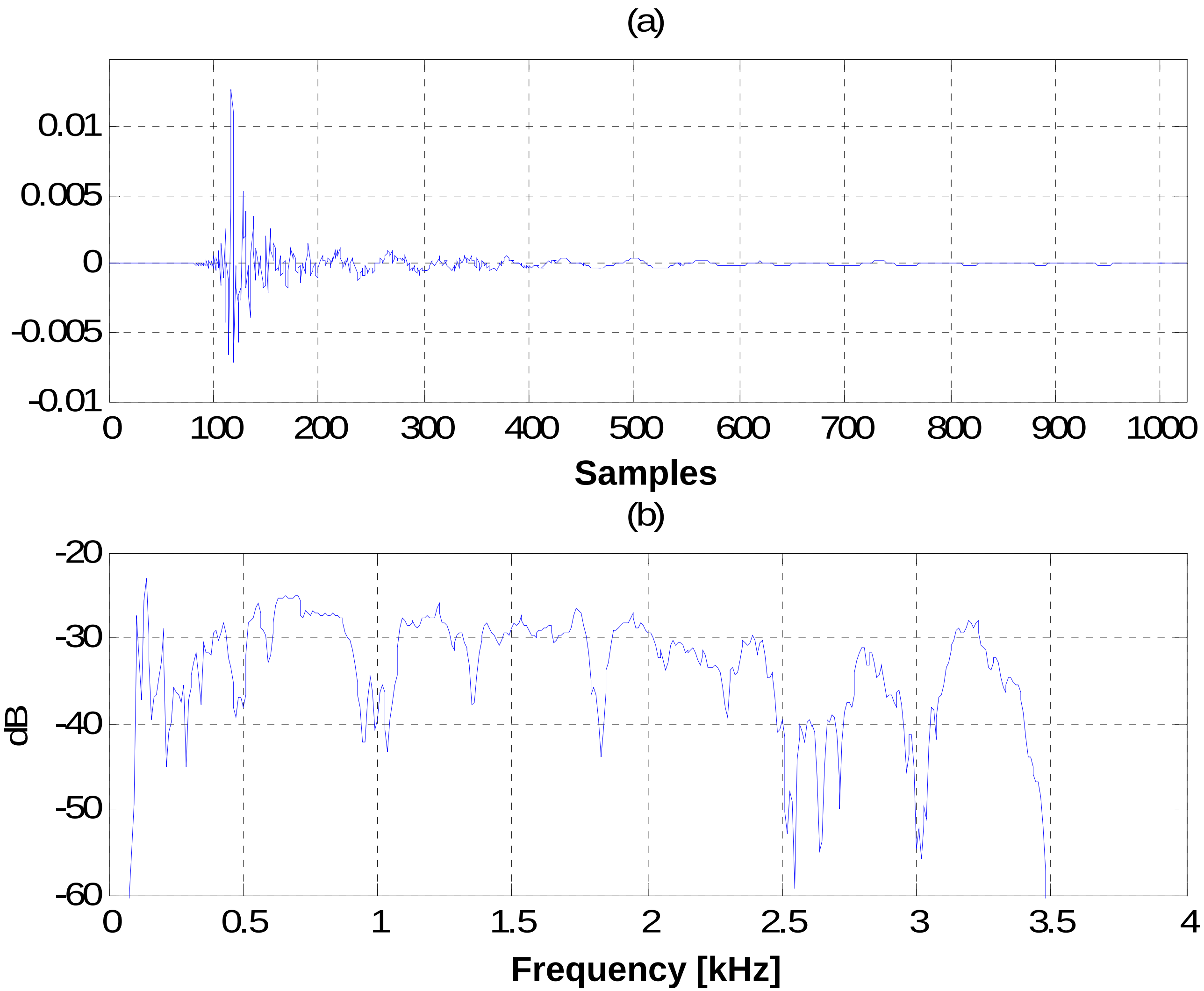


Fig. 1. Acoustic echo path: (a) impulse response; (b) frequency response.

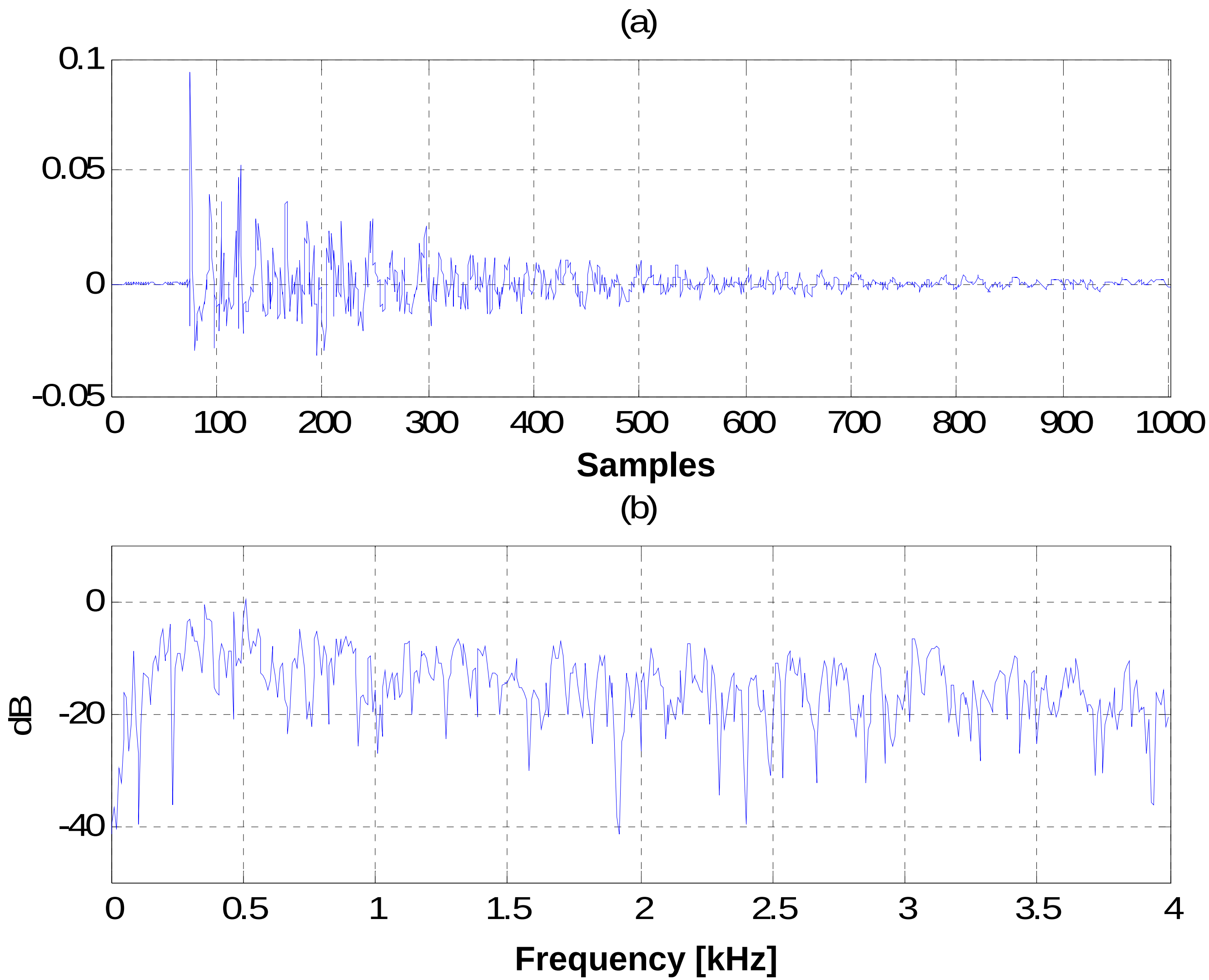


Fig. 2. Acoustic echo path: (a) impulse response; (b) frequency response.

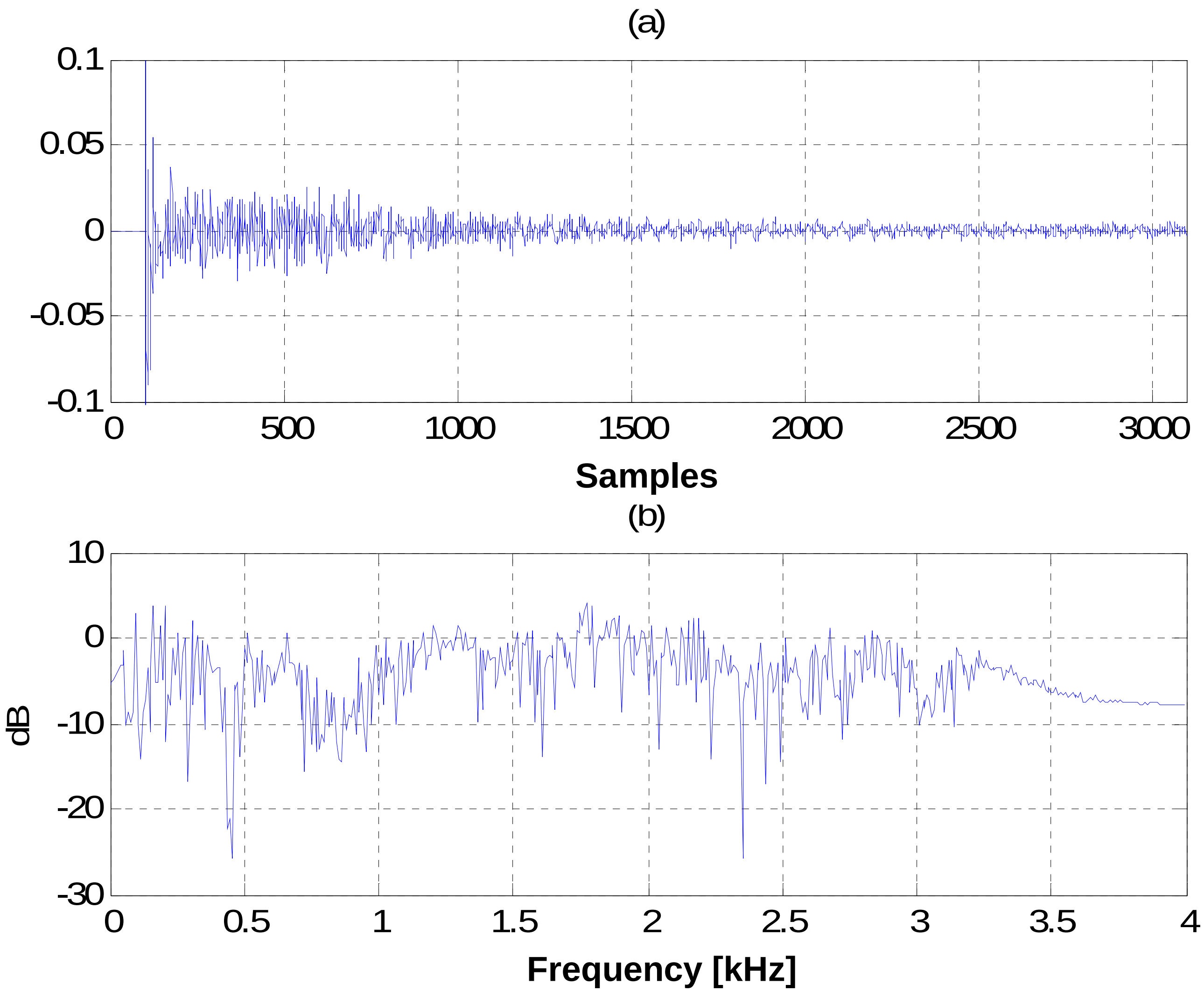


Fig. 3. Acoustic echo path: (a) impulse response; (b) frequency response.

Introduction

- acoustic echo cancellation (AEC)

  → required in ***hands-free*** communication devices, (e.g., for mobile telephony or teleconferencing systems)

  → **! acoustic coupling** between the loudspeaker and microphone

  → an ***adaptive filter*** identifies the acoustic echo path between the terminal's loudspeaker and microphone

- specific problems in AEC

  → the echo path can be extremely long

  → it may rapidly change at any time during the connection

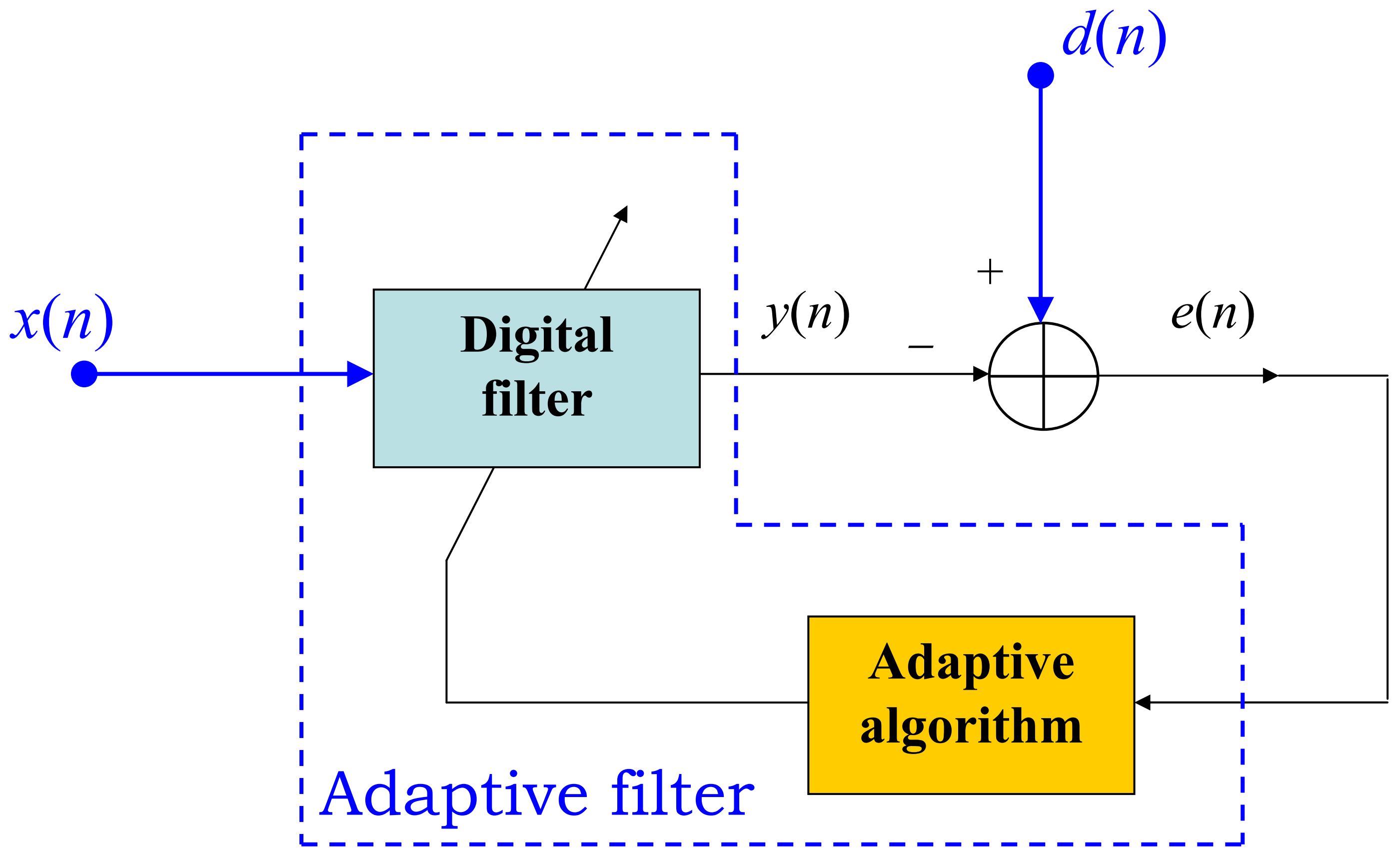
  → the background noise can be strong and non-stationary

- important issue in echo cancellation

  → the behaviour during double-talk

  → the presence of Double-Talk Detector (DTD)

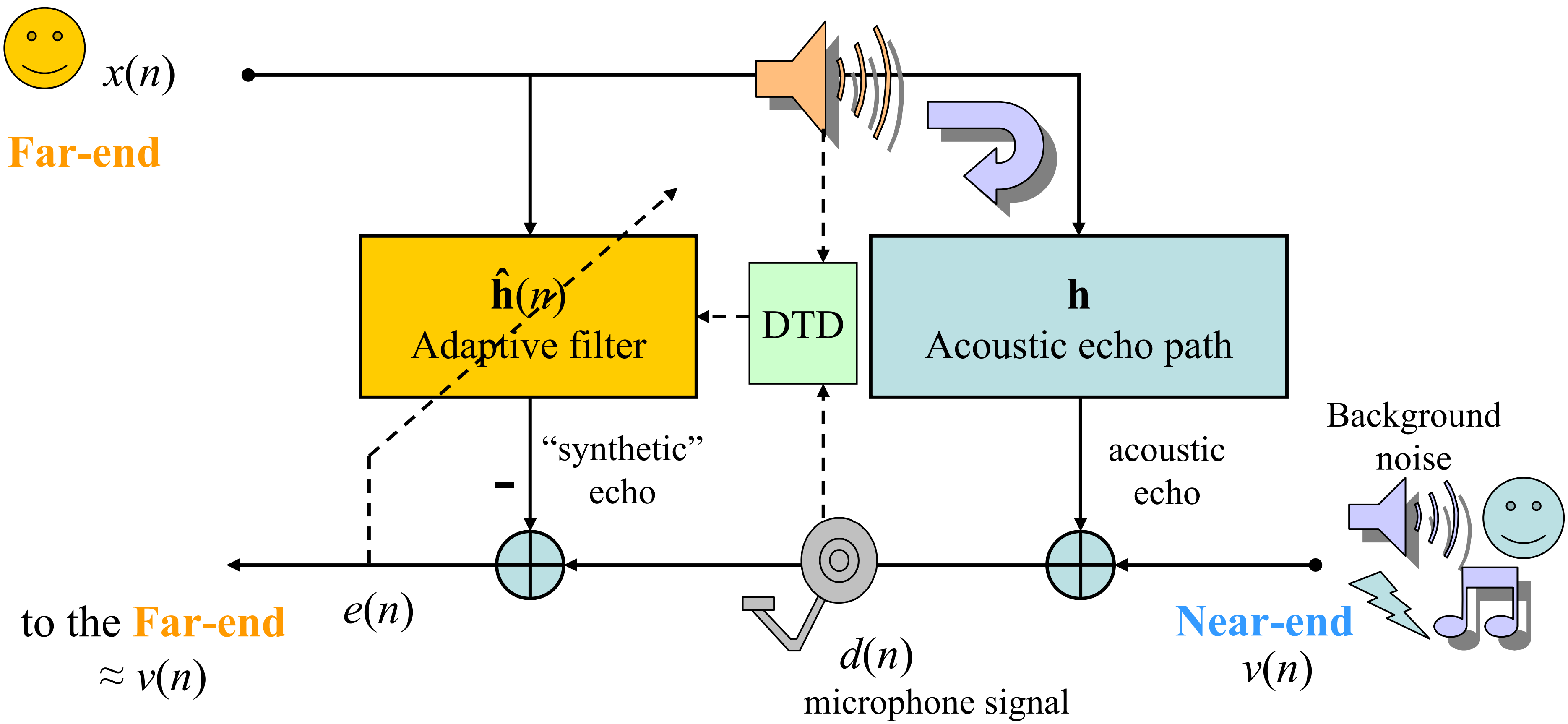
Introduction (cont.)



$e(n) = d(n) - y(n)$ → Cost function $J[e(n)] \downarrow$ minimized

Introduction (cont.)

- **AEC configuration**



Performance criteria: - convergence rate ***vs.*** misadjustment

- tracking ***vs.*** robustness

Adaptive algorithms for AEC

- **requirements**

  → fast convergence rate and tracking

  → low misadjustment

  → double-talk robustness

- **most common choices**

  → normalized least-mean-square (NLMS) algorithm

  → affine projection algorithm (APA)

step-size parameter

$\mathbf{0 < \mu \leq 1}$

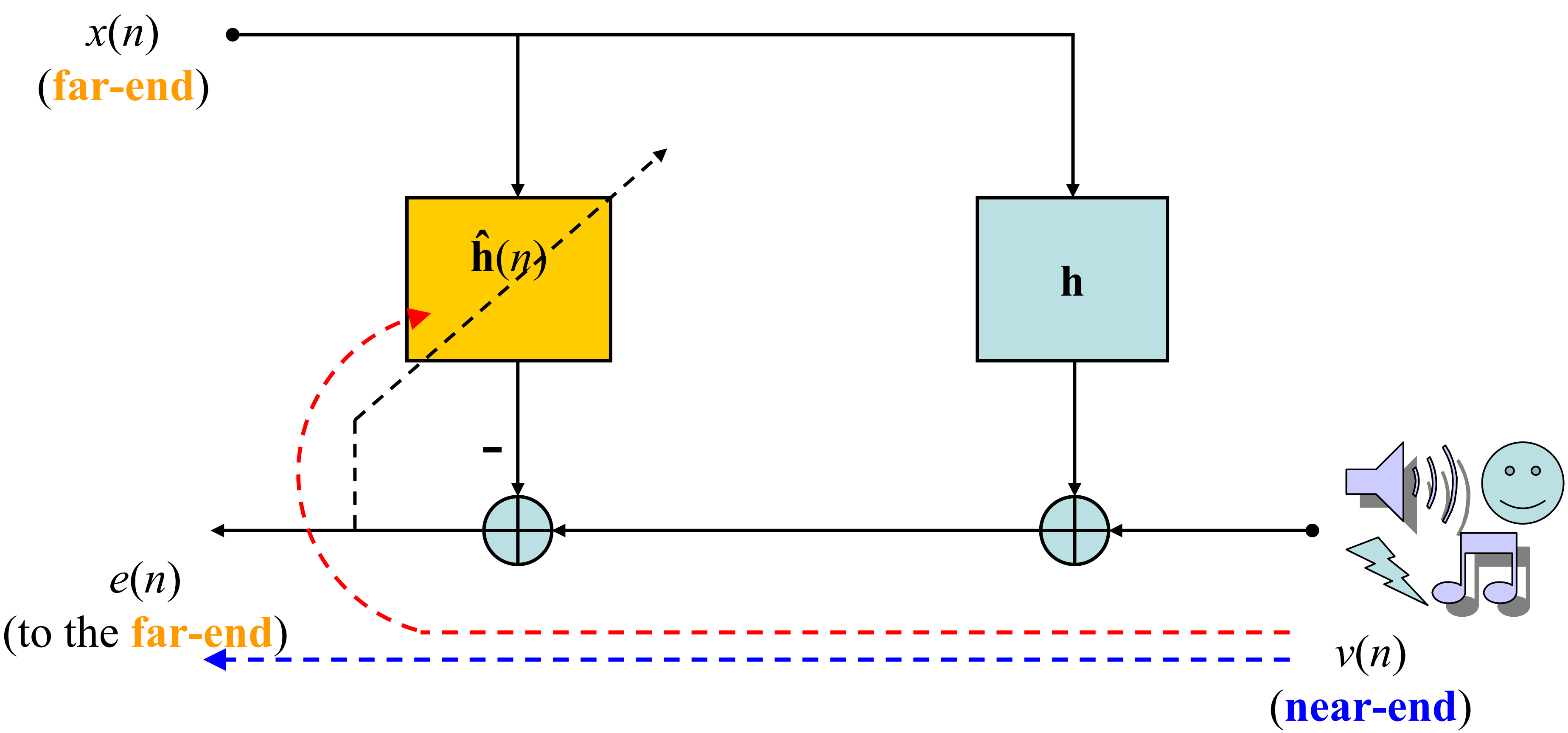
$$\boxed{\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mu \; \textbf{update-term}}$$

- **step-size parameter** (controls the performance of these algorithms)

  → *large* values → fast convergence rate and tracking

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

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



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

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

$$\boxed{\mu(n) = 1 - \frac{\hat{\sigma}_v(n)}{\hat{\sigma}_e(n)}}$$

?

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

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

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

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

$$\mu(n) = 1 - \frac{\hat{\sigma}_w}{\hat{\sigma}_e(n)}$$

background noise power estimate

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

➔ **NPVSS-NLMS** algorithm

Problem: background noise can be time-variant

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

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

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

near-end speech power estimate

???

Problem: non-stationary character of the speech signal

- **Solutions for evaluating the near-end signal power estimate**

$$\boxed{\hat{\sigma}_v^2(n) = ?}$$

1. using the error signal $e(n)$, with a larger value of the weighting factor:

$$\hat{\sigma}_e^2(n) = \lambda \hat{\sigma}_e^2(n-1) + (1-\lambda) e^2(n) \qquad \lambda = 1 - 1/(KL), \text{ with } K > 1$$

$$\boxed{\hat{\sigma}_v^2(n) = \gamma \hat{\sigma}_v^2(n-1) + (1-\gamma) e^2(n)} \qquad \gamma = 1 - 1/(QL), \text{ with } Q > K$$

➔ simple VSS-NLMS (**SVSS-NLMS**) algorithm

$$\boxed{\mu_{\text{SVSS}}(n) = 1 - \frac{\hat{\sigma}_v(n)}{\hat{\sigma}_e(n)}}$$

! The value of γ influences the overall behaviour of the algorithm.

2. using a normalized cross-correlation based echo path change detector:

$$\boxed{\hat{\sigma}_v^2(n) = \hat{\sigma}_e^2(n) - \frac{1}{\hat{\sigma}_x^2(n)} \hat{\mathbf{r}}_{e\mathbf{x}}^T(n) \hat{\mathbf{r}}_{e\mathbf{x}}(n)}$$

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

$$\hat{\mathbf{r}}_{e\mathbf{x}}(n) = \lambda \hat{\mathbf{r}}_{e\mathbf{x}}(n-1) + (1-\lambda) \mathbf{x}(n) e(n)$$

➔ **NEW-NPVSS-NLMS** algorithm

$$\boxed{\mu_{\text{NEW-NPVSS}}(n) = \begin{cases} 1 - \dfrac{\hat{\sigma}_v(n)}{\hat{\sigma}_e(n)} & \text{if } \xi(n) < \varsigma \\ 1 & \text{otherwise} \end{cases}}$$

[M. A. Iqbal *et al*, "Novel variable step size NLMS algorithms for echo cancellation," *Proc. IEEE ICASSP*, 2008]

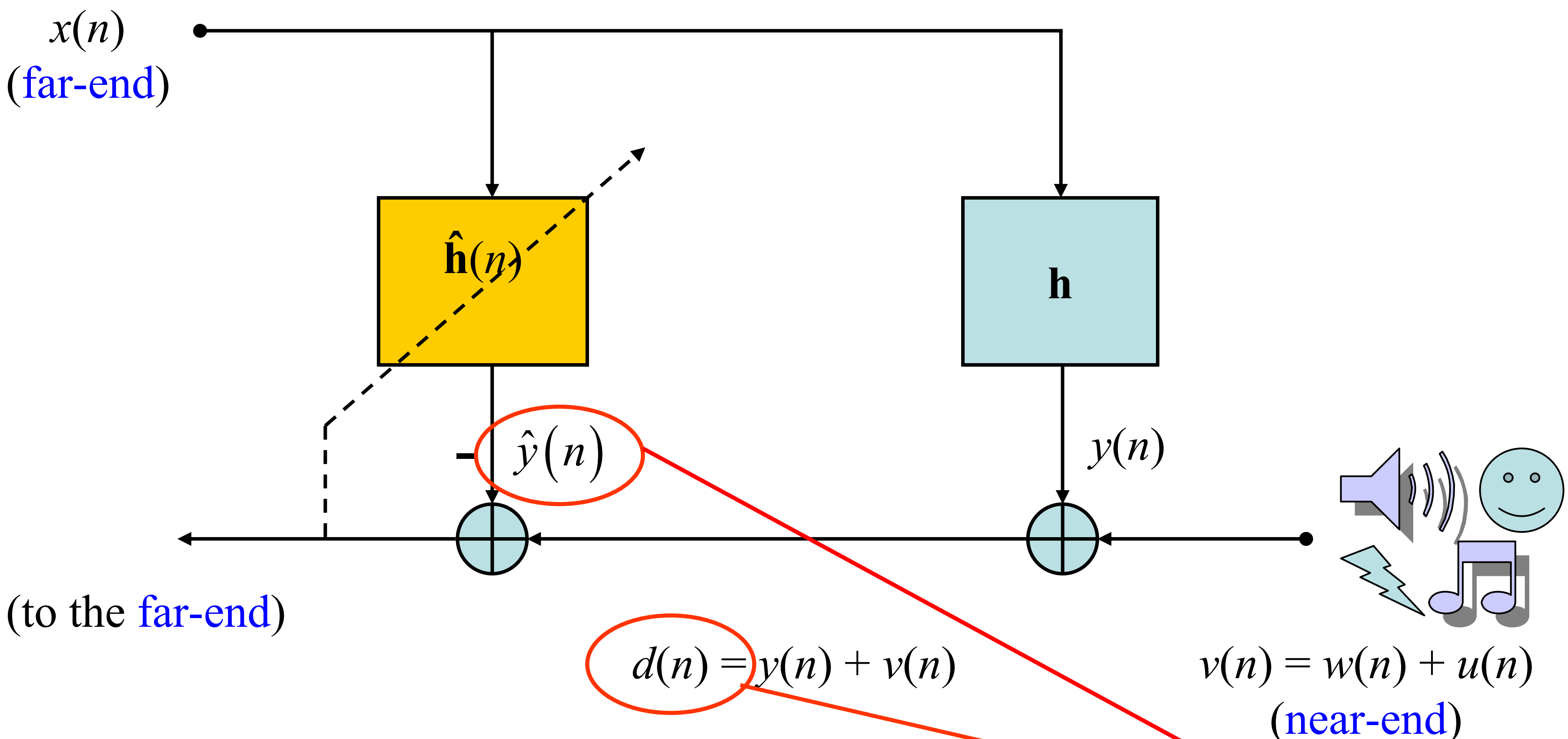
$$\xi(n) = \left| \frac{\hat{r}_{ed}(n) - \hat{\sigma}_e^2(n)}{\hat{\sigma}_d^2(n) - \hat{r}_{ed}(n)} \right|$$

convergence statistic

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

$$\hat{r}_{ed}(n) = \lambda \hat{r}_{ed}(n-1) + (1-\lambda) e(n) d(n)$$

! The value of ς influences the overall behaviour of the algorithm.



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

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

3.

$$\mu(n) = 1 - \frac{\sqrt{\hat{\sigma}_d^2(n) - \hat{\sigma}_{\hat{y}}^2(n)}}{\hat{\sigma}_e(n)}$$

➔ **Practical VSS-NLMS (PVSS-NLMS) algorithm**

[C. Paleologu, S. Ciochina, and J. Benesty, "Variable step-size NLMS algorithm for under-modeling acoustic echo cancellation," *IEEE Signal Process. Lett.*, 2008]

➔ **VSS affine projection algorithm (VSS-APA)**

[C. Paleologu, J. Benesty, and S. Ciochina, "Variable step-size affine projection algorithm designed for acoustic echo cancellation," *IEEE Trans. Audio, Speech, Language Process.*, Nov. 2008]

- **main advantages**

  → non-parametric algorithms

  → robustness to background noise variations and double-talk

! they assume that the adaptive filter has converged to a certain degree.

Table I. Computational complexities of the different variable-step sizes.

Algorithms	*Additions*	*Multiplications*	*Divisions*	*Square-roots*
NPVSS-NLMS	3	3	1	1
SVSS-NLMS	4	5	1	1
NEW-NPVSS-NLMS	$2L + 8$	$3L + 12$	3	1
PVSS-NLMS	6	9	1	1

L = adaptive filter length

Simulation results

- **conditions**

  → Acoustic echo cancellation (AEC) context, $L = 1000$.

  → input signal $x(n)$ – AR(1) signal or speech sequence.

  → background noise $w(n)$ – independent white Gaussian noise signal (variable SNR)

  → measure of performance – normalized misalignment (dB)

$$20\log_{10}(\| \mathbf{h} - \hat{\mathbf{h}}(n) \| / \| \mathbf{h} \|)$$

- **algorithms for comparisons**

  → NLMS

  → NPVSS-NLMS

  → SVSS-NLMS

  → NEW-NPVSS-NLMS

  → PVSS-NLMS

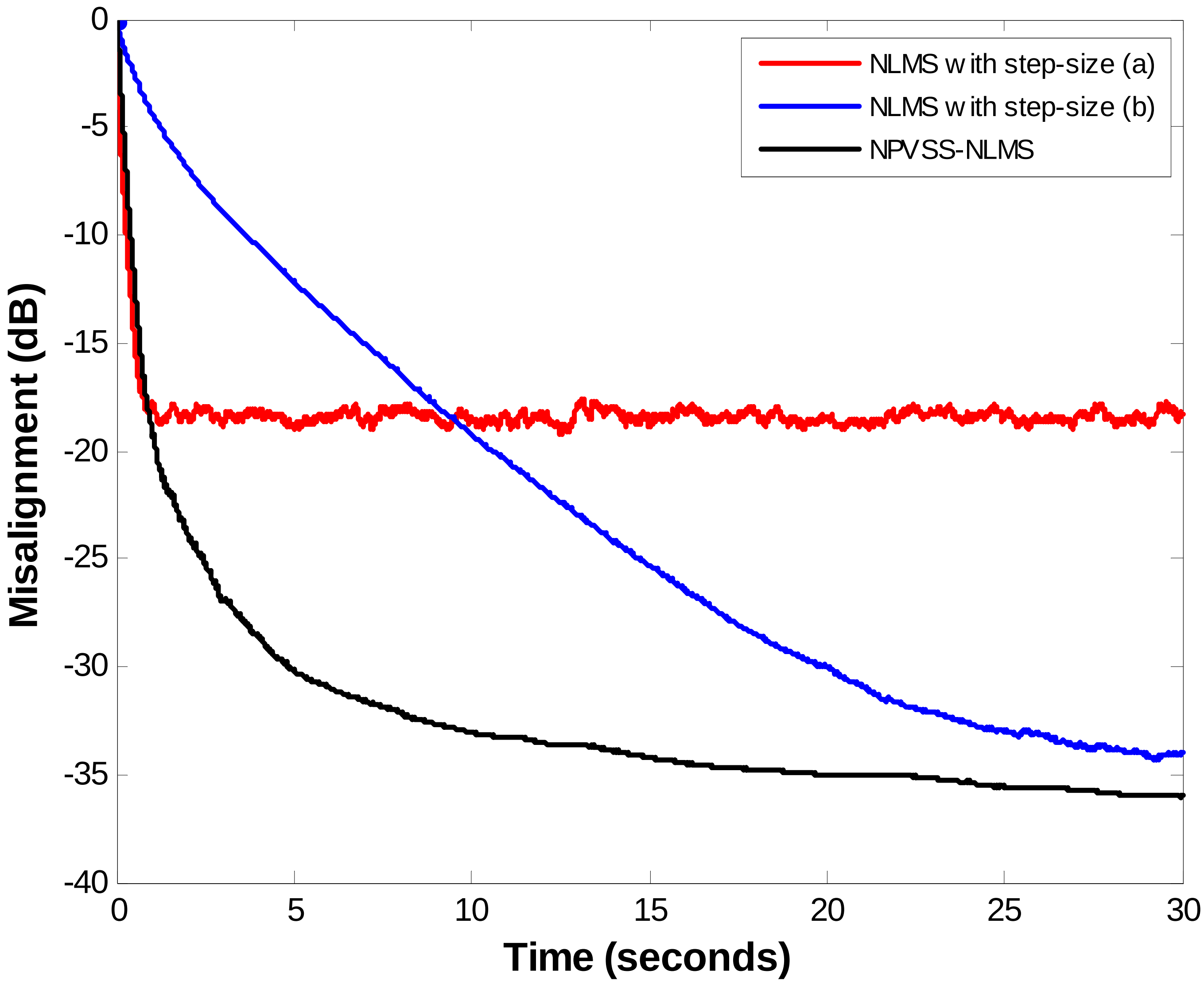


Fig. 4. Misalignment of the NLMS algorithm with two different step sizes (a) $\mu = 1$ and (b) $\mu = 0.05$, and misalignment of the NPVSS-NLMS algorithm. The input signal is an AR(1) process, $L = 1000$, $\lambda = 1 - 1/(6L)$, and SNR = 20 dB.

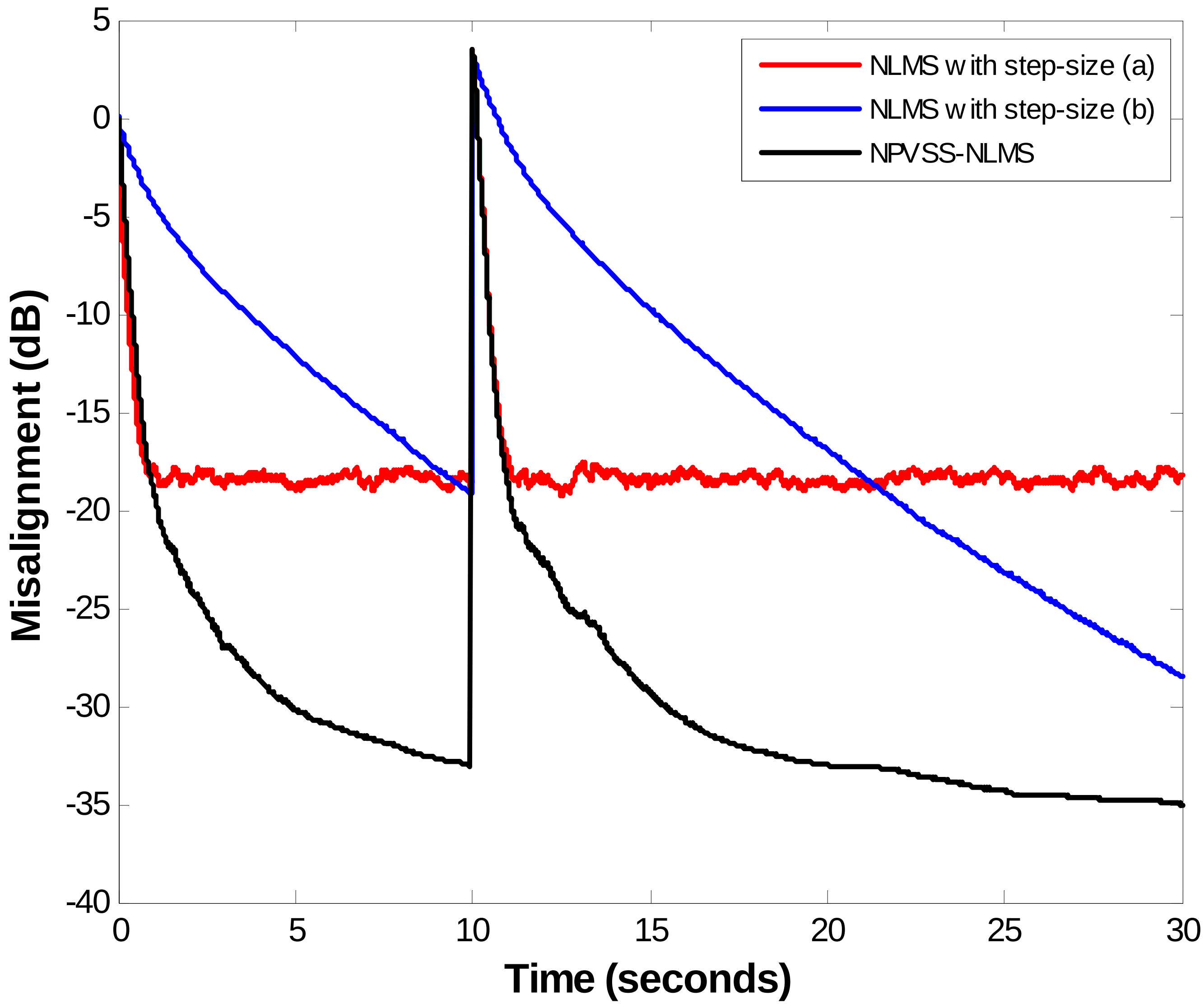


Fig. 5. Misalignment of the NLMS algorithm with two different step sizes (a) $\mu = 1$ and (b) $\mu = 0.05$, and misalignment of the NPVSS-NLMS algorithm. The input signal is an AR(1) process, $L = 1000$, $\lambda = 1 - 1/(6L)$, and SNR = 20 dB. Echo path changes at time 10.

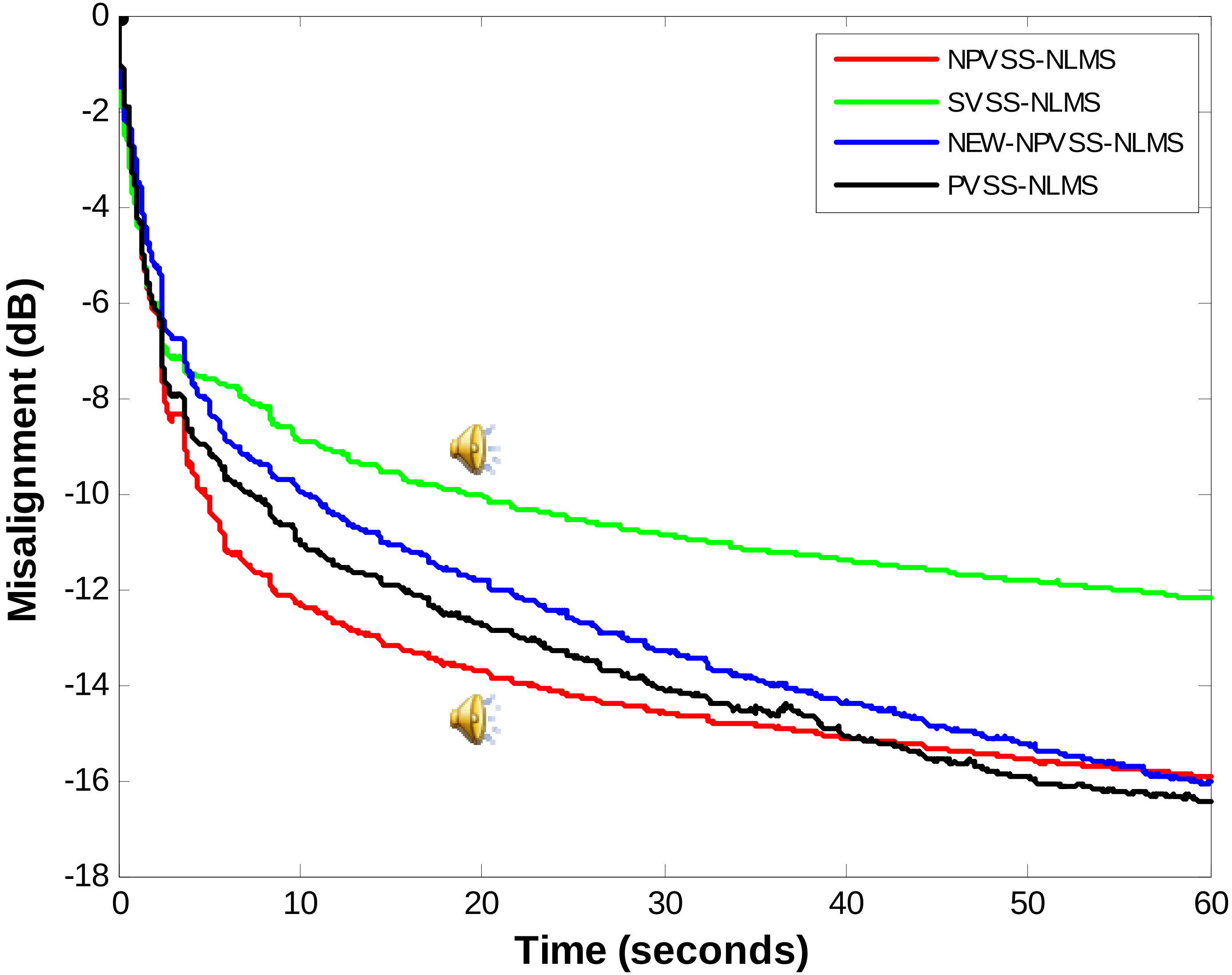


Fig. 6. Misalignment of the NPVSS-NLMS, SVSS-NLMS [with $\gamma = 1 - 1/(18L)$], NEW-NPVSS-NLMS (with $\varsigma = 0.1$), and PVSS-NLMS algorithms. The input signal is speech, $L = 1000$, $\lambda = 1 - 1/(6L)$, and SNR = 20 dB.

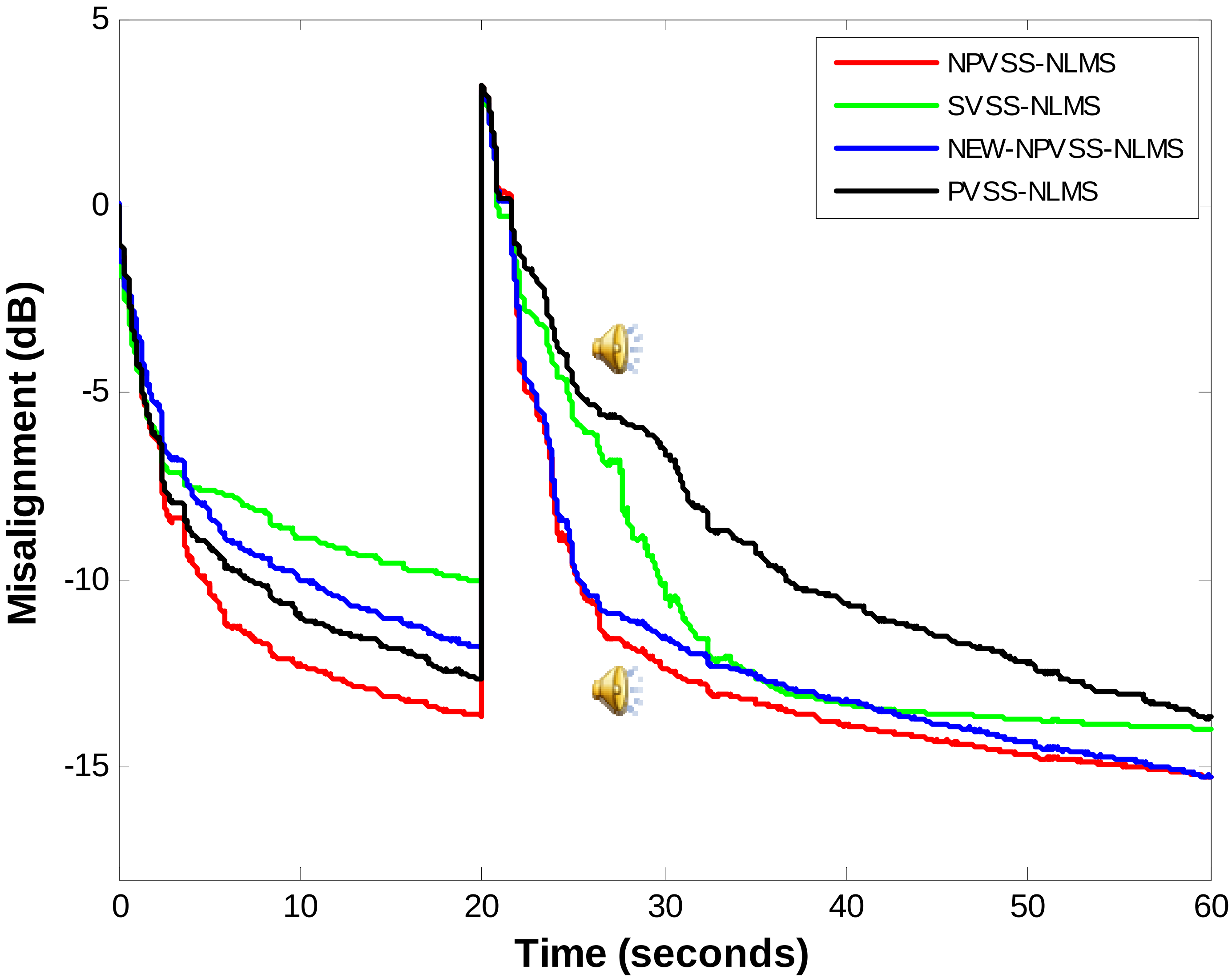


Fig. 7. Misalignment during impulse response change. The impulse response changes at time 20. Algorithms: NPVSS-NLMS, SVSS-NLMS [with $\gamma = 1 - 1/(18L)$], NEW-NPVSS-NLMS (with $\varsigma = 0.1$), and PVSS-NLMS. The input signal is speech, $L = 1000$, $\lambda = 1 - 1/(6L)$, and SNR = 20 dB.

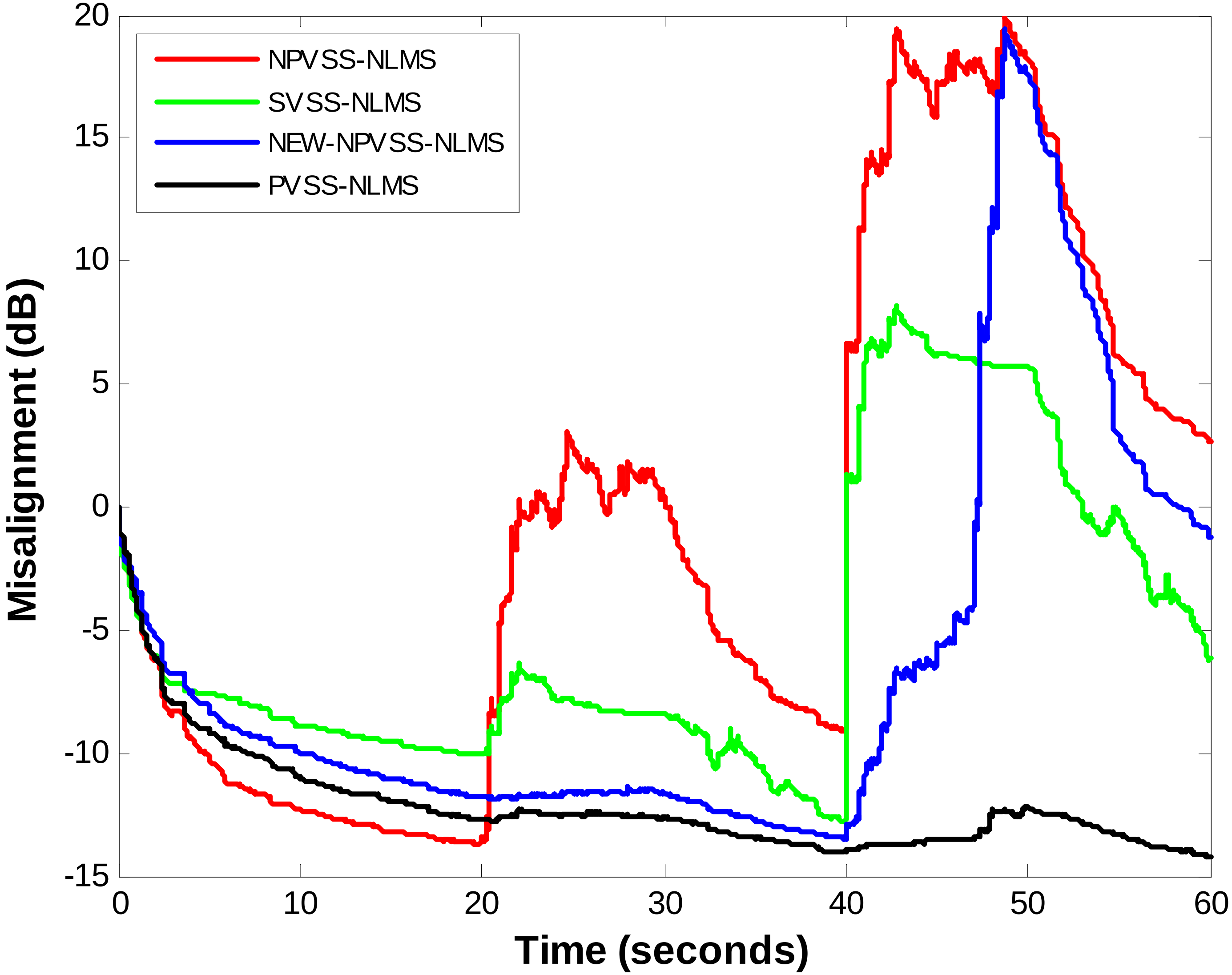


Fig. 8. Misalignment during background noise variations. The SNR decreases from 20 dB to 10 dB between time 20 and 30, and to 0 dB between time 40 and 50. Algorithms: NPVSS-NLMS, SVSS-NLMS [with $\gamma = 1 - 1/(18L)$], NEW-NPVSS-NLMS (with $\varsigma = 0.1$), and PVSS-NLMS. The input signal is speech, $L = 1000$, $\lambda = 1 - 1/(6L)$, and SNR = 20 dB.

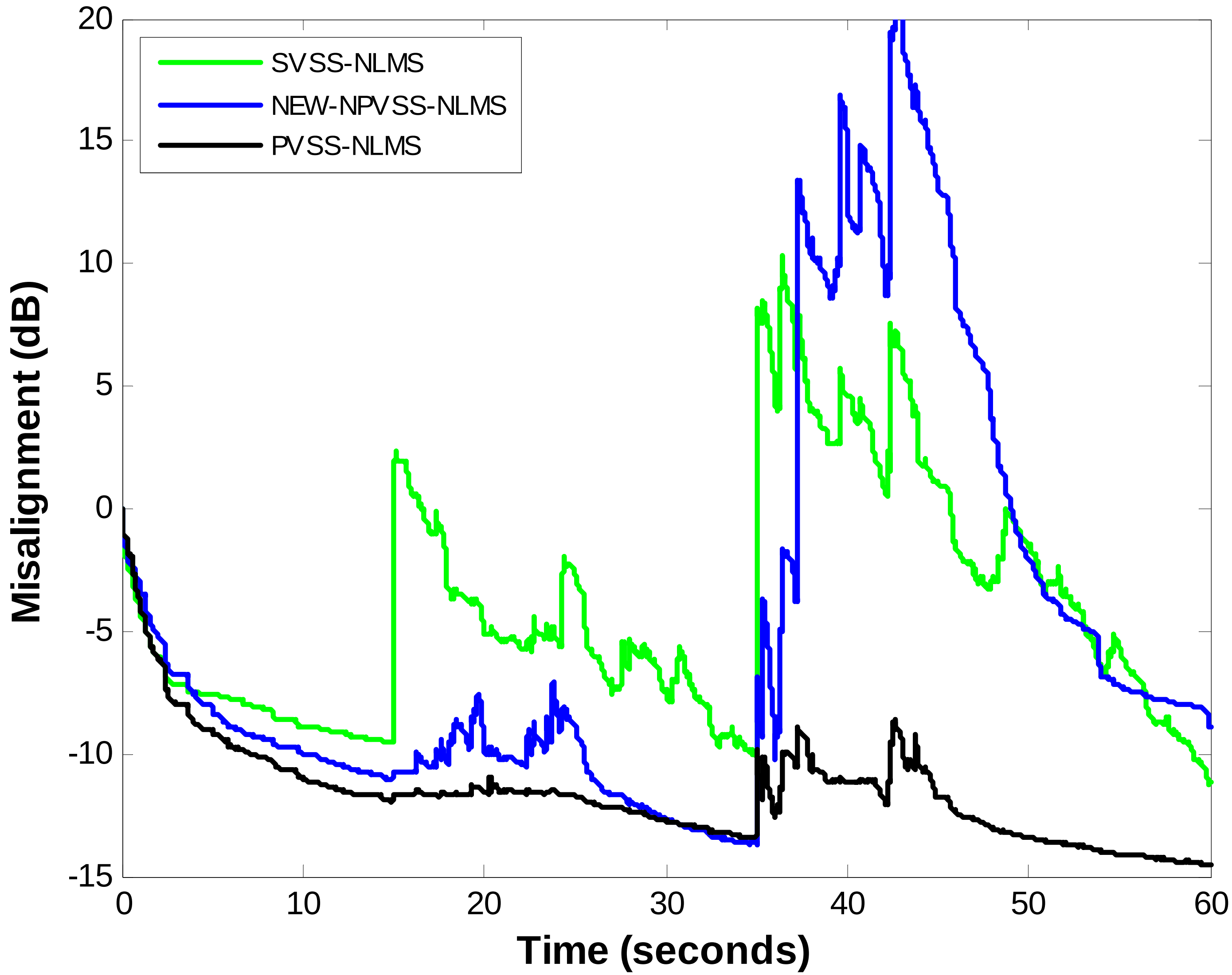


Fig. 9. Misalignment during double-talk, without DTD. Near-end speech appears between time 15 and 25 (with FNR = 5 dB), and between time 35 and 45 (with FNR = 3 dB). Algorithms: SVSS-NLMS [with $\gamma = 1 - 1/(18L)$], NEW-NPVSS-NLMS (with $\varsigma = 0.1$), and PVSS-NLMS. The input signal is speech, $L = 1000$, $\lambda = 1 - 1/(6L)$, and SNR = 20 dB.

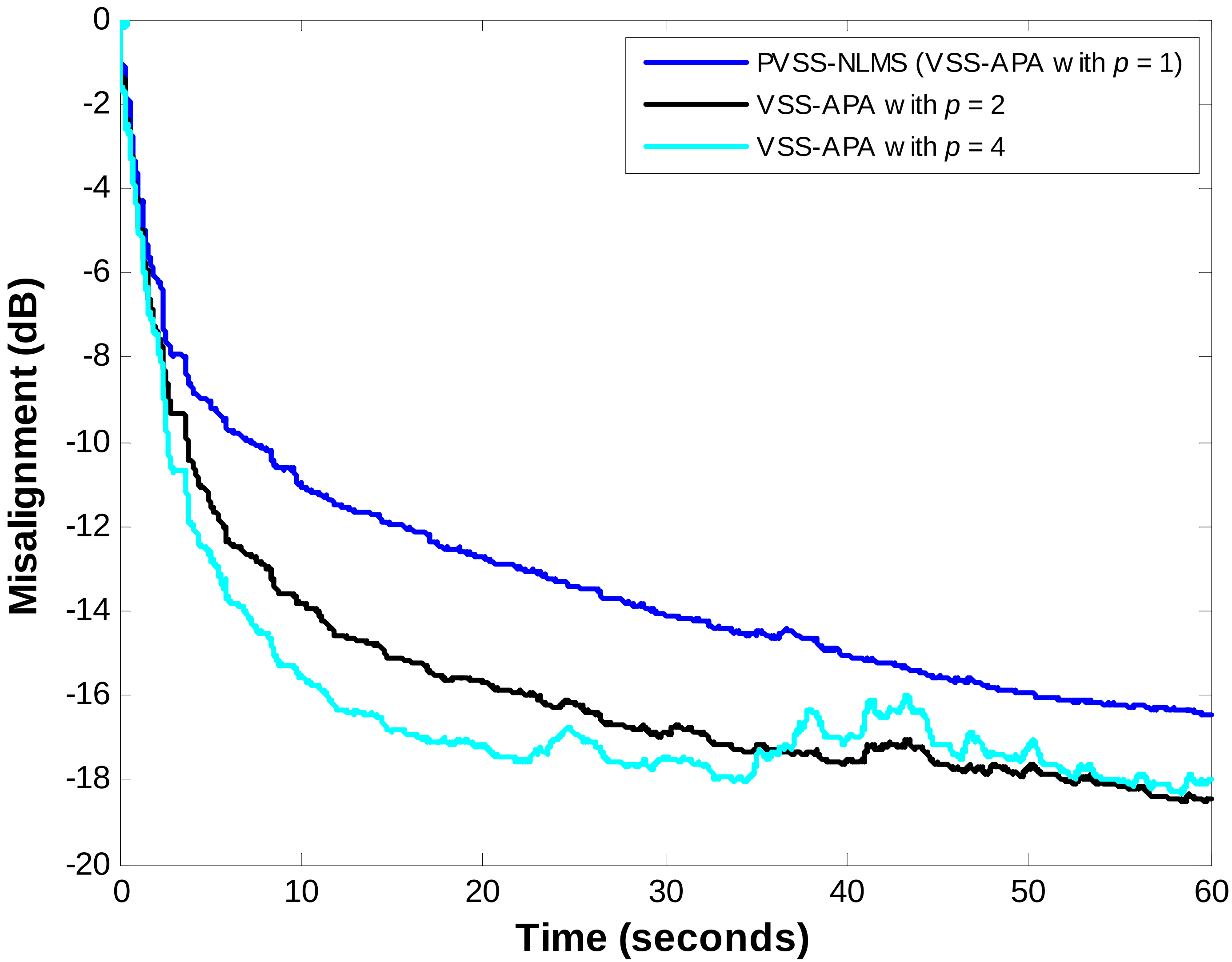


Fig. 10. Misalignment of the VSS-APA with different projection orders, i.e., $p = 1$ (PVSS-NLMS algorithm), $p = 2$, and $p = 4$. Other conditions are the same as in Fig. 12.

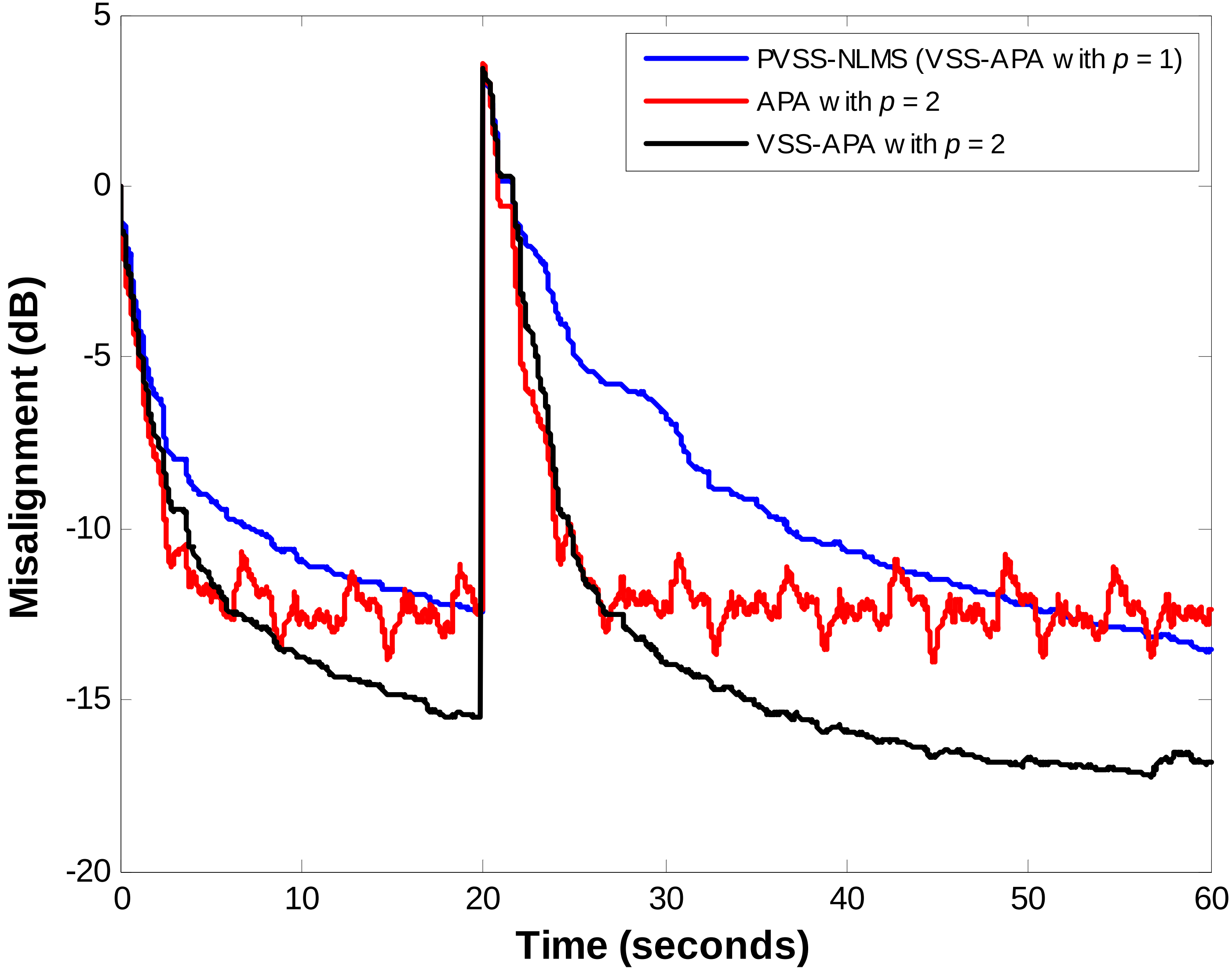


Fig. 11. Misalignment during impulse response change. The impulse response changes at time 20. Algorithms: PVSS-NLMS algorithm, APA (with $\mu = 0.25$), and VSS-APA. Other conditions are the same as in Fig. 12.

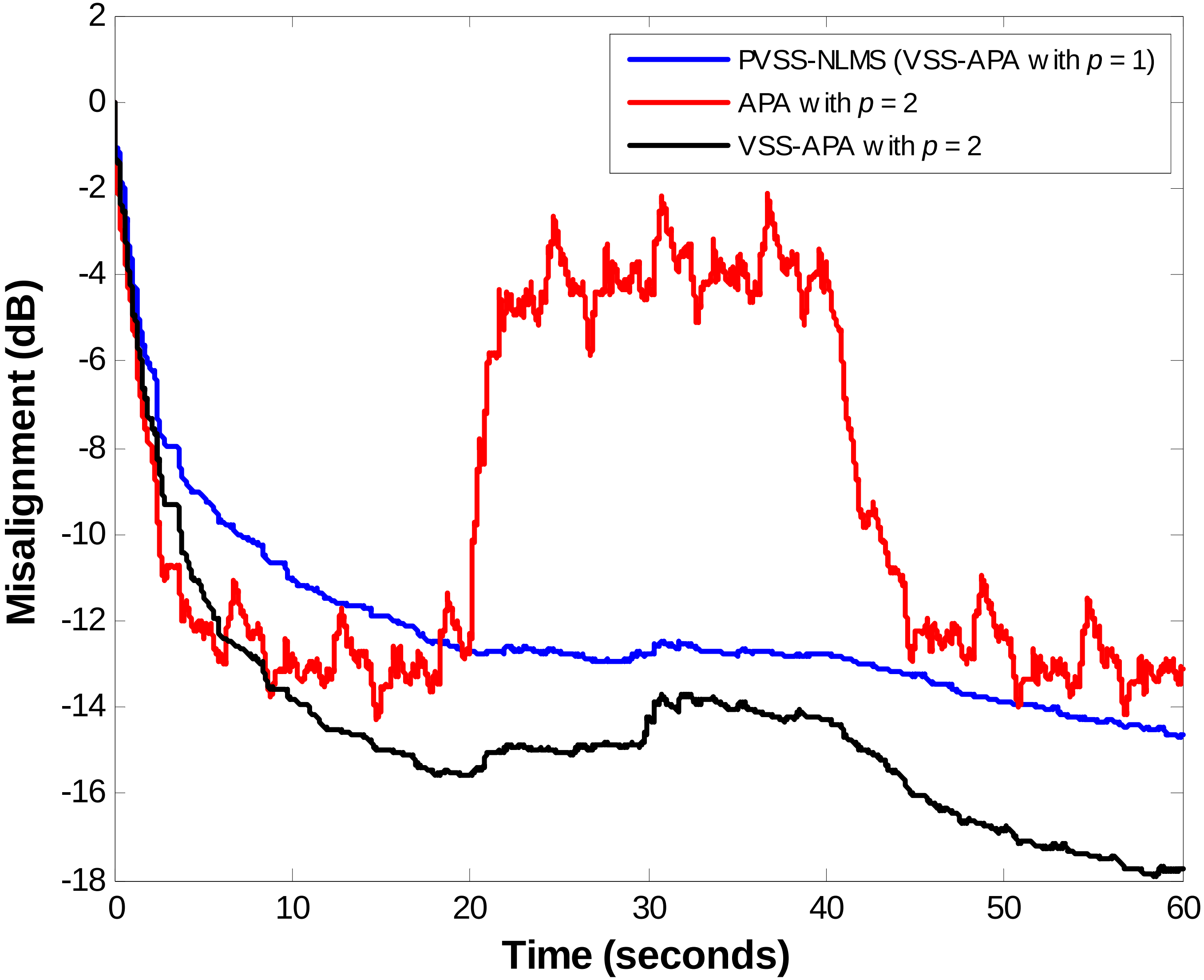


Fig. 12. Misalignment during background noise variations. The SNR decreases from 20 dB to 10 dB between time 20 and 40. Other conditions are the same as in Fig. 12.

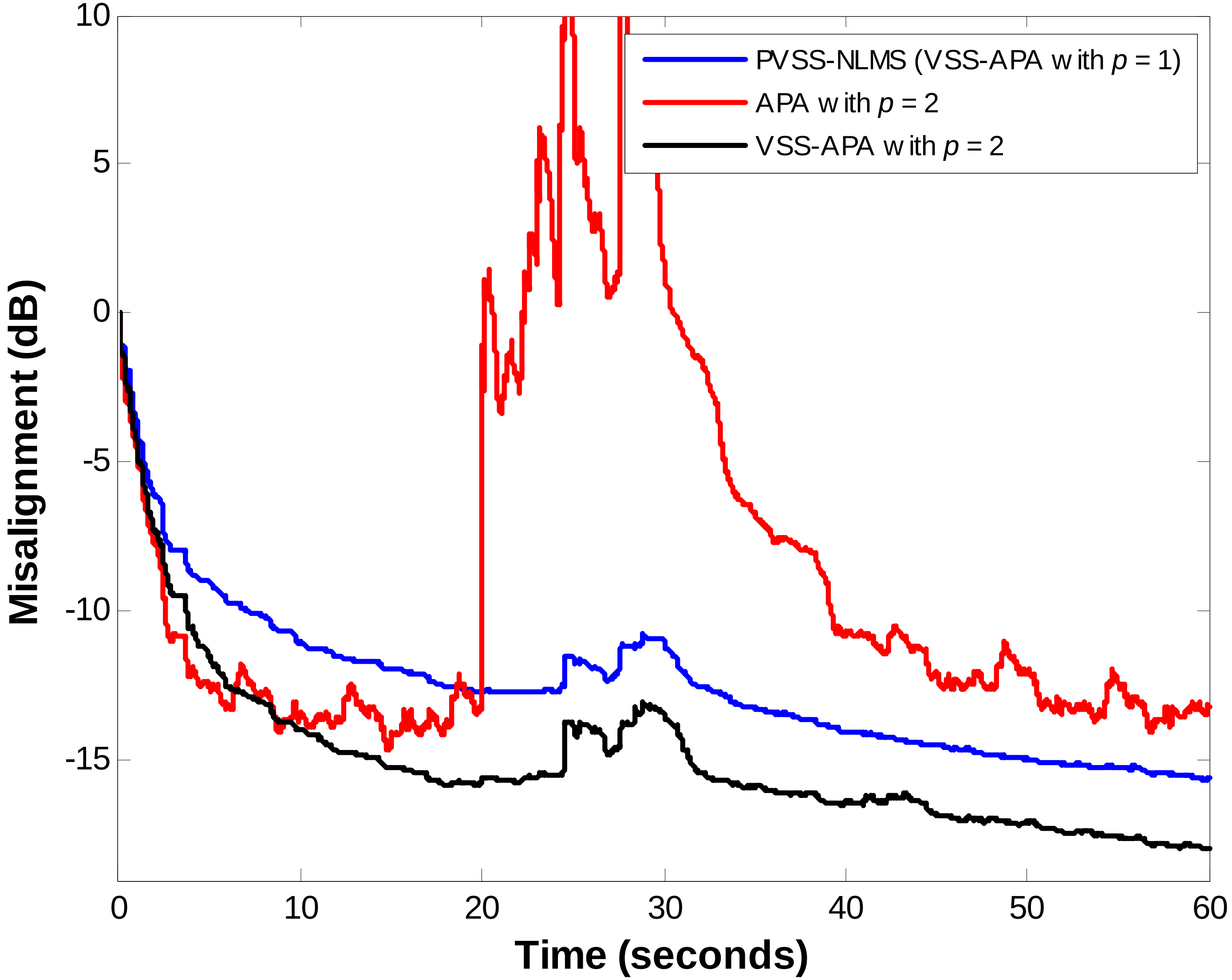


Fig. 13. Misalignment during double-talk, without a DTD. Near-end speech appears between time 20 and 30 (with FNR = 4 dB). Other conditions are the same as in Fig. 12.

Comparisons with other VSS-type APAs

- **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

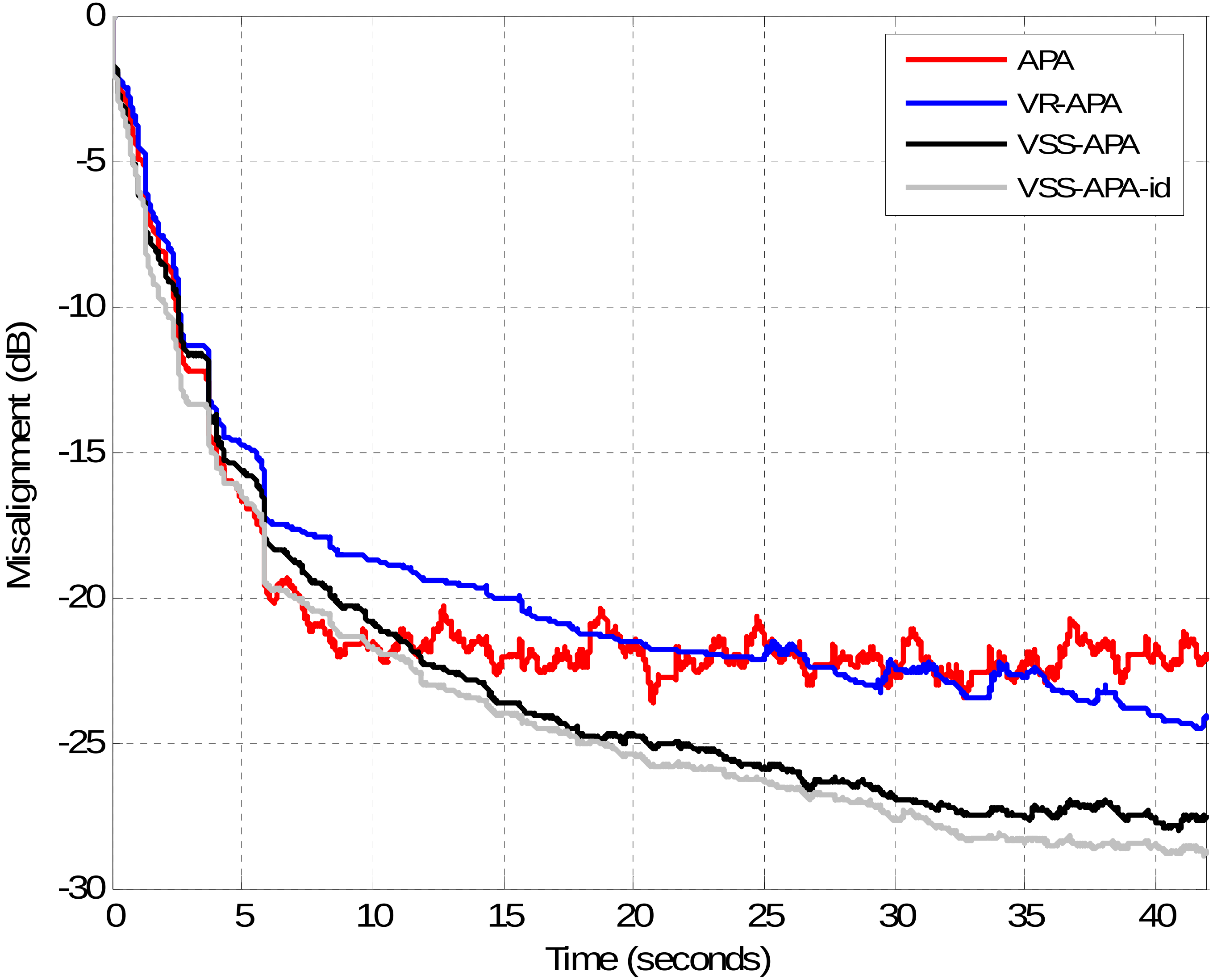


Fig. 14. Misalignments of APA with $\mu = 0.2$, VR-APA, VSS-APA, and VSS-APA-id. Single-talk case, $L = 512$, $p = 2$ for all the algorithms, SNR = 20dB.

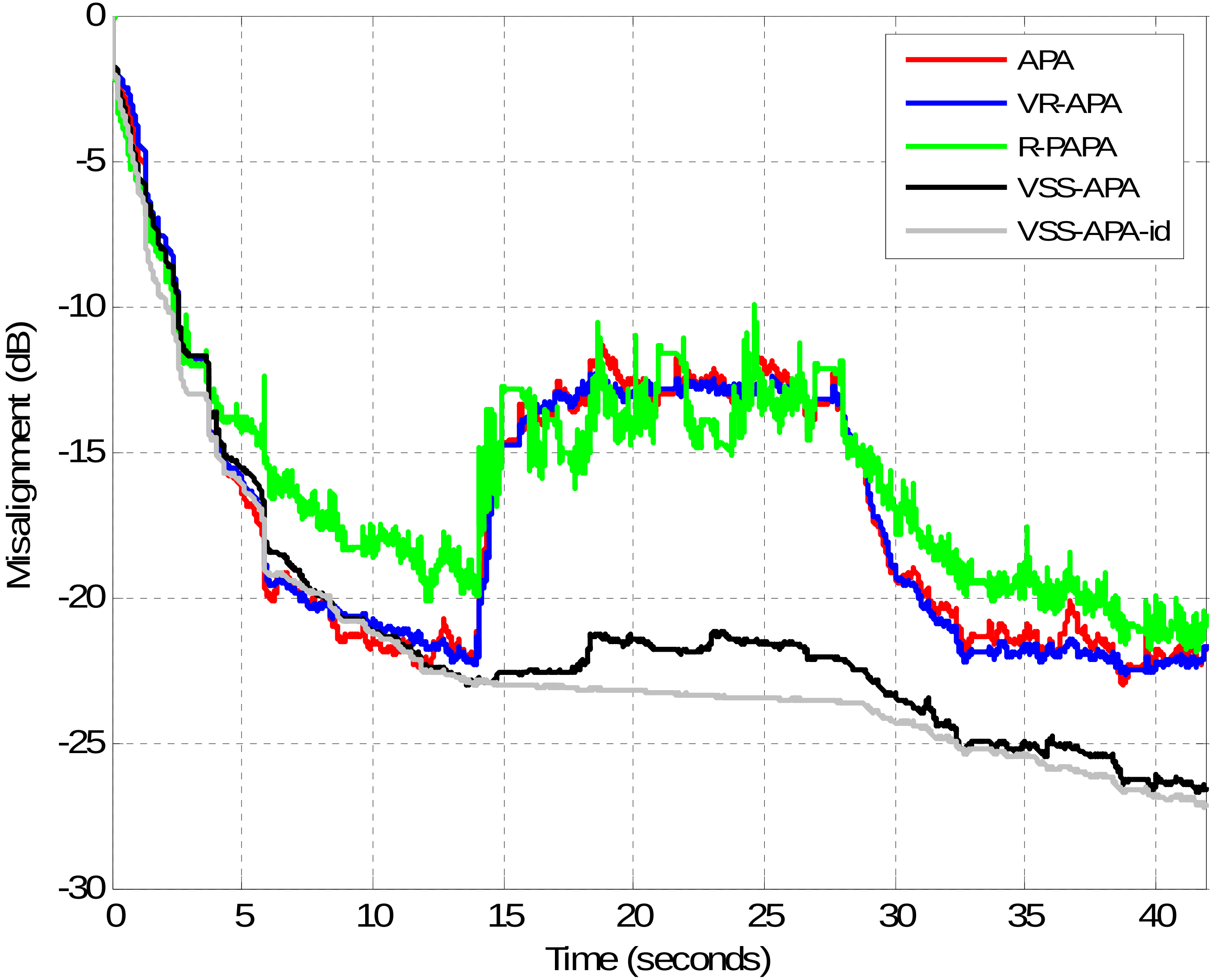


Fig. 15. Misalignments of APA, VR-APA, R-PAPA, VSS-APA, and VSS-APA-id. Background noise variation at time 14, for a period of 14 seconds (SNR decreases from 20dB to 10 dB). Other conditions are the same as in Fig. 14.

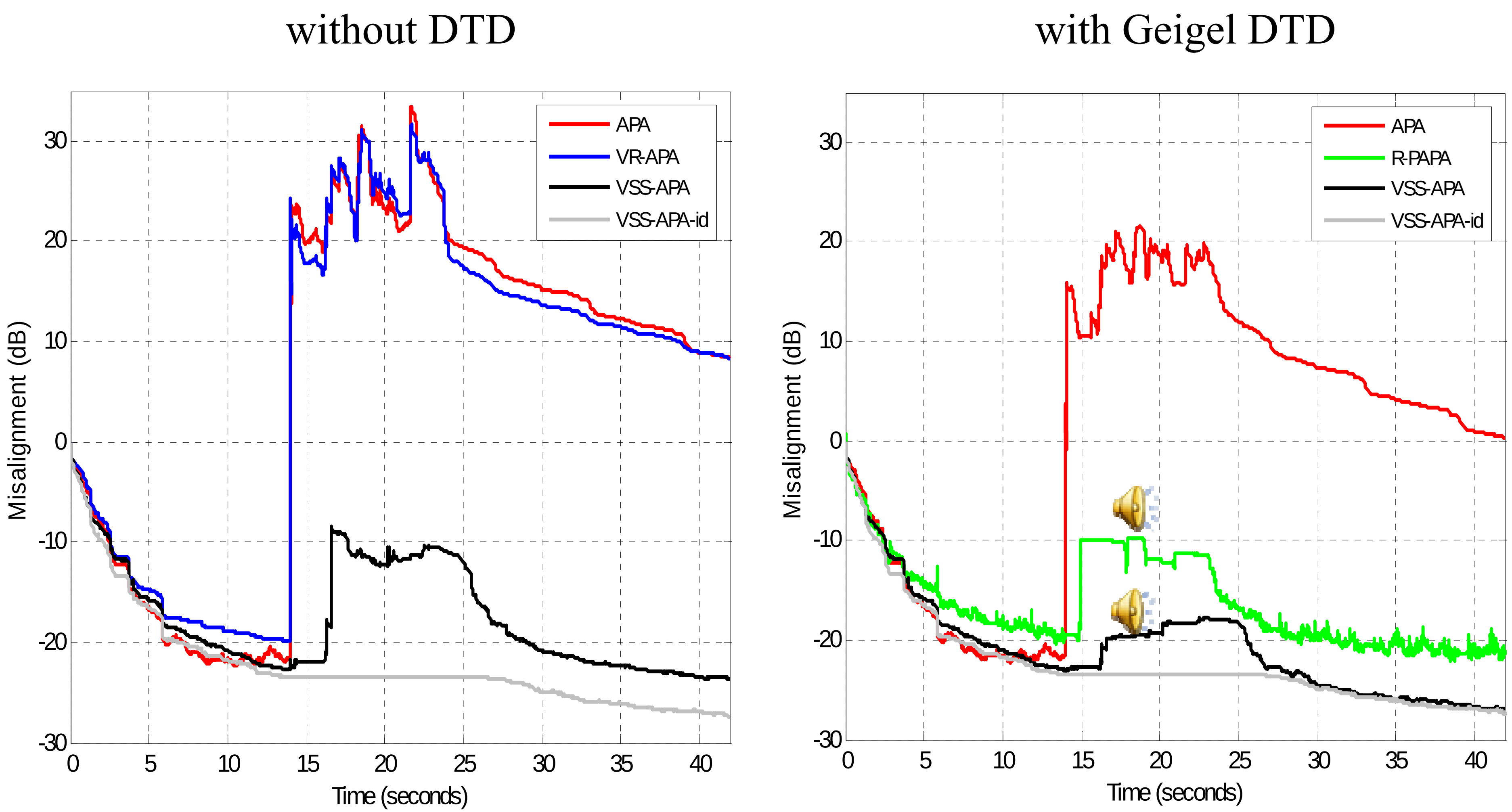


Fig. 16. Misalignment of the algorithm during double-talk. Other conditions are the same as in Fig. 14.

Conclusions and Perspectives

- a family of VSS-type algorithms was developed in the context of AEC.
- the VSS formulas do 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.
- the experimental results indicate that these algorithms are reliable candidates for real-world applications.

[C. Anghel, C. Paleologu, *et al*,"FPGA implementation of a variable step-size affine projection algorithm for acoustic echo cancellation," in *Proc. EUSIPCO,* 2010]

[C. Stanciu, C. Anghel, C. Paleologu, *et al*,"FPGA implementation of an efficient proportionate affine projection algorithm for echo cancellation," in *Proc. EUSIPCO,* 2011]

Perspectives

- Future work ➔ towards proportionate adaptive algorithms
  ➔ towards stereophonic AEC

• **[C. Paleologu, J. Benesty, and S. Ciochină, *Sparse Adaptive Filters for Echo Cancellation*, Morgan & Claypool Publishers, ISBN 978-1-598-29306-7, 2010]**

• **[C. Paleologu, S. Ciochină, and J. Benesty, "An efficient proportionate affine projection algorithm for echo cancellation," *IEEE Signal Processing Letters*, 2010]**

• **[J. Benesty, C. Paleologu, and S. Ciochină, "Proportionate adaptive filters from a basis pursuit perspective," *IEEE Signal Processing Letters*, 2010]**

• **[C. Paleologu, J. Benesty, F. Albu, and S. Ciochină, "An efficient variable step-size proportionate affine projection algorithm," in *Proc. IEEE ICASSP*, 2011]**

• **[J. Benesty, C. Paleologu, T. Gänsler, and S. Ciochină, *A Perspective on Stereophonic Acoustic Echo Cancellation*, Springer-Verlag, ISBN 978-3-642-22573-4, 2011]**

Thank you!