16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

ROBUST VARIABLE STEP-SIZE AFFINE PROJECTION ALGORITHM SUITABLE FOR ACOUSTIC ECHO CANCELLATION

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

[1] Department of Telecommunications, University Politehnica of Bucharest
1-3, Iuliu Maniu Blvd., 061071, Bucharest, Romania
phone: + (40) 21 4024765, fax: + (40) 21 4024765, email: {pale, silviu}@comm.pub.ro
[2] INRS-EMT, Universite du Quebec
QC H5A 1K6, Montreal, Canada
phone: + (1) 514 875-1226/ext. 2019, fax: + (1) 514 875-0344, email: benesty@emt.inrs.ca

ABSTRACT

The affine projection algorithm (APA) and different versions of it have proved to be very attractive choices for acoustic echo cancellation (AEC). In this context, a classical APA with a constant step-size has to compromise between two performance criteria, i.e., 1) high convergence rates and good tracking capabilities, and 2) low misadjustment and robustness against background noise variations and double-talk. Consequently, a variable step-size APA (VSS-APA) is a more reliable solution. In this paper we propose a VSS-APA that is designed to recover the near-end signal from the error signal of the adaptive filter. Therefore, it is robust against near-end signal variations, including double-talk. Moreover, since it does not require a priori information about the acoustic environment, the proposed algorithm is easy to control in real-world AEC applications.

1. INTRODUCTION

In acoustic echo cancellation (AEC) applications, an adaptive filter identifies the acoustic echo path between the terminal's loudspeaker and microphone, i.e., the room impulse response [1]. In this context, several specific problems have to be addressed. First, the echo path can be extremely long and it may rapidly change at any time during the connection. Secondly, the background noise that corrupts the microphone signal can be strong and non-stationary in nature. Third, the involved signals (i.e., mainly speech) are non-stationary and highly correlated. Finally, the behaviour during double-talk (i.e., the talkers on both sides speak simultaneously) has to be considered.

The affine projection algorithm (APA) [2] and some of its versions, e.g., [3], [4], were found to be very attractive choices for AEC applications, since they offer a superior convergence rate as compared to the normalized least-mean-square (NLMS) algorithm, especially for speech signals. The performance of the classical APA is governed by the step-size parameter. This parameter has to be chosen based on a compromise between fast convergence rate and good tracking capabilities on the one hand, and low misadjustment on the other hand. In order to meet this conflicting requirement, a number of variable step-size APAs (VSS-APAs) were developed [5], [6] (and reference therein). Unfortunately, most of the existing VSS-APAs require the tuning of some parameters (e.g., background noise power) which are not a priori available or have to be estimated.

The overall performance of any adaptive filter used in AEC could be seriously affected during double-talk, up to divergence. In this case, the standard procedure is to use a double-talk detector (DTD) in order to slow down or completely halt the adaptation process during double-talk periods. Nevertheless, there is some inherent delay in the decision of a DTD; during this small period a few undetected large amplitude samples can perturb the echo path estimate considerably. Consequently, it is highly desirable to improve the robustness of the adaptive algorithm in order to handle a certain amount of double-talk without diverging. This is the motivation behind the development of the so-called robust algorithms [1].

In this paper, we propose a VSS-APA suitable for AEC. Most of the APAs aim to cancel p (i.e., projection order) previous a posteriori errors at every step of the algorithm. The proposed approach takes into account the fact that the microphone signal contains the background noise or/and a speech sequence, and these signals should be recovered in the error signal of the adaptive filter. The resulting formula of the step-size depends only on signals that are available within the AEC application, i.e., the output signal of the adaptive filter and the microphone signal. Thus, there is no need for a priori information about the acoustic environment, so that the algorithm is non-parametric from this point of view. Due to the basis of its derivation, the proposed VSS-APA is robust to near-end signal variations like background noise increase or double-talk.

The paper is organized as follows. Section 2 introduces the classical APA, followed by the derivation of the proposed VSS-APA. The simulation results are presented in Section 3. Finally, Section 4 concludes this work.

2. PROPOSED VSS-APA FOR AEC

The goal of a general AEC configuration is to identify an acoustic echo path $\mathbf{h} = [h_0\ h_1\ \dots\ h_{L-1}]^T$, using an adaptive filter $\hat{\mathbf{h}}(n) = [\hat{h}_0(n)\ \hat{h}_1(n)\ \dots\ \hat{h}_{L-1}(n)]^T$, where superscript T denotes transposition and n is the time index; we assume that both systems are real-valued finite impulse responses of

This work was supported by the UEFISCSU Romania under Grants PN-II-"Idei" no. 65/01.10.2007 and no. 331/01.10.2007.

length L. The far-end speech $x(n)$ goes through the acoustic impulse response $\mathbf{h}$, resulting the echo signal, $y(n)$. This signal is picked up by the microphone together with the near-end signal $v(n)$, resulting the microphone signal $d(n)$. The near-end signal can contain both the background noise, $w(n)$, and the near-end speech, $u(n)$. The output of the adaptive filter, $\hat{y}(n)$, provides a replica of the echo, which will be subtracted from the microphone signal.

The APA [2] is defined by the following relations:

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

$$\hat{\mathbf{h}}(n)=\hat{\mathbf{h}}(n-1)+\mu\mathbf{X}(n)\left[\mathbf{X}^T(n)\mathbf{X}(n)\right]^{-1}\mathbf{e}(n), \quad (2)$$

where $\mathbf{e}(n)$ is the a priori error vector and $\mathbf{d}(n) = [d(n), d(n-1), \ldots, d(n-p+1)]^T$ is the desired signal vector of length p, with p denoting the projection order. The matrix $\mathbf{X}(n) = [\mathbf{x}(n), \mathbf{x}(n-1), \ldots, \mathbf{x}(n-p+1)]$ is the input signal matrix, where $\mathbf{x}(n-l) = [x(n-l), x(n-l-1), \ldots, x(n-l-L+1)]^T$ (with $l=0,1,\ldots,p-1$) are the input signal vectors. The constant μ denotes the step-size parameter of the algorithm.

Equation (2) can be rewritten in a different form as

$$\hat{\mathbf{h}}(n)=\hat{\mathbf{h}}(n-1)+\mathbf{X}(n)\left[\mathbf{X}^T(n)\mathbf{X}(n)\right]^{-1}\boldsymbol{\mu}(n)\mathbf{e}(n), \quad (3)$$

where

$$\boldsymbol{\mu}(n)=\text{diag}\left\{\mu_0(n),\mu_1(n),\ldots,\mu_{p-1}(n)\right\} \quad (4)$$

is a $p \times p$ diagonal matrix. We can recover (2) imposing that $\mu_0(n)=\mu_1(n)=\ldots=\mu_{p-1}(n)=\mu$.

The a posteriori error vector can be defined using the adaptive filter coefficients at time n:

$$\boldsymbol{\varepsilon}(n)=\mathbf{d}(n)-\mathbf{X}^T(n)\hat{\mathbf{h}}(n). \quad (5)$$

Replacing (3) in (5) and taking (1) into account, it results that

$$\boldsymbol{\varepsilon}(n)=\left[\mathbf{I}_p-\boldsymbol{\mu}(n)\right]\mathbf{e}(n), \quad (6)$$

where $\mathbf{I}_p$ denotes a $p \times p$ identity matrix. The basic idea of the APA imposes to cancel p a posteriori errors, i.e., $\boldsymbol{\varepsilon}(n)=\mathbf{0}_{p\times 1}$, where $\mathbf{0}_{p\times 1}$ denotes a column vector with all its p elements equal to zeros. Assuming that $\mathbf{e}(n)\neq\mathbf{0}_{p\times 1}$, it results from (6) that $\boldsymbol{\mu}(n)=\mathbf{I}_p$. This corresponds to the classical APA update from (2), with the step-size $\mu = 1$. In the absence of the near-end signal (i.e., $v(n) = 0$, which leads to an ideal "system identification" configuration) the value of the step-size $\mu = 1$ makes sense, because it leads to the best performance [2].

Nevertheless, the AEC scheme can be viewed as an "interference cancelling" configuration, aiming to recover an "useful" signal (i.e., the near-end signal) corrupted by an undesired perturbation (i.e., the acoustic echo); consequently, the "useful" signal should be recovered in the error signal of the adaptive filter. Therefore, a more reasonable condition is $\boldsymbol{\varepsilon}(n) = \mathbf{v}(n)$, where $\mathbf{v}(n) = [v(n), v(n-1), \ldots, v(n-p+1)]^T$ represents the near-end signal vector of length p. Taking (6) into account, it results that

$$\varepsilon_{l+1}(n)=\left[1-\mu_l(n)\right]e_{l+1}(n)=v(n-l), \quad (7)$$

where $\varepsilon_{l+1}(n)$ and $e_{l+1}(n)$ denote the (l+1)-th elements of the vectors $\boldsymbol{\varepsilon}(n)$ and $\mathbf{e}(n)$, with $l=0,1,\ldots,p-1$. The expression of the step-size parameter $\mu_l(n)$ has to be found such that

$$E\left\{\varepsilon_{l+1}^2(n)\right\}=E\left\{v^2(n-l)\right\}, \quad (8)$$

where $E\{\bullet\}$ denotes mathematical expectation. Squaring (7) and taking the expectations it results:

$$\left[1-\mu_l(n)\right]^2E\left\{e_{l+1}^2(n)\right\}=E\left\{v^2(n-l)\right\}. \quad (9)$$

By solving the quadratic equation (9), we obtain

$$\mu_l(n)=1-\sqrt{\frac{E\left\{v^2(n-l)\right\}}{E\left\{e_{l+1}^2(n)\right\}}}. \quad (11)$$

From a practical point of view, (11) has to be evaluated in terms of power estimates as

$$\mu_l(n)=1-\frac{\hat{\sigma}_v(n-l)}{\hat{\sigma}_{e_{l+1}}(n)}. \quad (12)$$

The denominator from (12) can be computed in a recursive manner, i.e.,

$$\hat{\sigma}_{e_{l+1}}^2(n)=\lambda\hat{\sigma}_{e_{l+1}}^2(n-1)+(1-\lambda)e_{l+1}^2(n), \quad (13)$$

where λ is a weighting factor chosen as $\lambda = 1 - 1/(KL)$, with $K > 1$; the initial value is $\hat{\sigma}_{e_{l+1}}^2(0)=0$.

The estimation of $\hat{\sigma}_v(n-l)$ is not straightforward in real-world applications like AEC. In this case, $v(n)$ is the near-end signal (i.e., background noise or/and near-end speech) which is combined together with the acoustic echo, resulting the microphone signal (the only signal that is practically available). Two main scenarios could be considered, as follows.

1) In the absence of the near-end speech, the near-end signal consists only of the background noise, $w(n)$. Its power can be estimated and it could be assumed constant, so that (12) becomes

$$\mu_l(n)=1-\frac{\hat{\sigma}_w}{\hat{\sigma}_{e_{l+1}}(n)}. \quad (14)$$

For a value of the projection order $p = 1$, the non-parametric VSS-NLMS (NPVSS-NLMS) algorithm proposed in [7] is obtained. For $p > 1$, a VSS-APA can be derived, by computing (14) for $l=0,1,\ldots,p-1$, then using a step-size matrix like in (4), and updating the filter coefficients according to (3). Nevertheless, the background noise can be time-variant, so that the power of the background noise should be periodically estimated. Moreover, when the background noise changes between two consecutive estimations or during the near-end speech, its new power estimate will not be available immediately; consequently, until the next estimation period of the background noise, the algorithm behaviour will be disturbed.

2) In the double-talk case, the near-end signal $v(n)$ consists of both the background noise, $w(n)$, and the near-end speech, $u(n)$; so that, $v(n) = w(n) + u(n)$. It is very difficult to obtain

an accurate estimate for the power of this combined signal, taking into account especially the non-stationary character of the speech signal. Therefore, (12) is futile and the presence of a double-talk detector is a must, in order to control the adaptation process during these periods.

Let us consider the previous cases in a more unified framework. The microphone signal at time index n can be expressed as

$$d(n) = y(n) + v(n), \tag{15}$$

where $y(n) = \mathbf{x}^T(n)\mathbf{h}$. Squaring (15) and taking the expectation of both sides [assuming that $y(n)$ and $v(n)$ are uncorrelated] it results that $E\{d^2(n)\} = E\{y^2(n)\} + E\{v^2(n)\}$. Assuming that the adaptive filter has converged to a certain degree, it can be considered that

$$E\{y^2(n)\} \cong E\{\hat{y}^2(n)\}, \tag{16}$$

where $\hat{y}(n) = \mathbf{x}^T(n)\hat{\mathbf{h}}(n-1)$. Consequently,

$$E\{v^2(n)\} \cong E\{d^2(n)\} - E\{\hat{y}^2(n)\}, \tag{17}$$

or in terms of power estimates [similar to (13)]

$$\hat{\sigma}_v^2(n) \cong \hat{\sigma}_d^2(n) - \hat{\sigma}_{\hat{y}}^2(n). \tag{18}$$

For the case 1), when only the background noise is present, i.e., $v(n) = w(n)$, an estimate of its power is obtained using the right-hand term in (18). This expression holds even if the level of the background noise changes, so that there is no need for the estimation of this parameter during silences. For the case 2), when the near-end speech is present (assuming that it is uncorrelated with the background noise), the near-end signal power estimate can be expressed as $\hat{\sigma}_v^2(n) = \hat{\sigma}_w^2(n) + \hat{\sigma}_u^2(n)$; the last parameter denotes the power estimate of the near-end speech. Accordingly, the right-hand term in (18) provides a power estimate of the near-end signal. Most importantly, this term depends only on the signals that are available within the AEC application, i.e., the microphone signal, $d(n)$, and the output of the adaptive filter, $\hat{y}(n)$. Consequently, (12) can be rewritten as

$$\mu_l(n) = 1 - \frac{\sqrt{\hat{\sigma}_d^2(n-l) - \hat{\sigma}_{\hat{y}}^2(n-l)}}{\hat{\sigma}_{e_{l+1}}(n)}, \tag{19}$$

for $l = 0, 1, \ldots, p-1$, which is more suitable in practice.

The adaptive filter coefficients should be updated using (3), with the step-sizes computed according to (19). In practice, (3) has to be rewritten as

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mathbf{X}(n)\left[\delta\mathbf{I}_p + \mathbf{X}^T(n)\mathbf{X}(n)\right]^{-1}\boldsymbol{\mu}(n)\mathbf{e}(n), \tag{20}$$

where δ is a positive scalar known as the regularization factor. An insightful analysis about this factor, in the framework of APA, can be found in [6].

Finally, some practical issues have to be addressed. First, a very small positive number ξ should be added to the denominator in (19) to avoid division by zero. Secondly, under our assumptions, we have $E\{d^2(n-l)\} \geq E\{\hat{y}^2(n-l)\}$ and $E\{d^2(n-l)\} - E\{\hat{y}^2(n-l)\} \cong E\{e^2_{l+1}(n)\}$. Nevertheless, the power estimates of these parameters could lead to some deviations from the previous theoretical conditions, so that we will take the absolute values in (19). Hence, the final step-sizes formula is written as

$$\mu_l(n) = \left|1 - \frac{\sqrt{\left|\hat{\sigma}_d^2(n-l) - \hat{\sigma}_{\hat{y}}^2(n-l)\right|}}{\xi + \hat{\sigma}_{e_{l+1}}(n)}\right|, \tag{21}$$

for $l = 0, 1, \ldots, p-1$. For a value of the projection order $p = 1$, the VSS-NLMS algorithm proposed in [8] is obtained.

Since it is based on the assumption that the adaptive filter coefficients have converged to a certain degree, the proposed VSS-APA could experienced a slower initial convergence rate and a slower tracking capability as compared to the APA, because (16) is biased in these situations. Nevertheless, the experimental results will prove that the performance degradation is not very significant (especially when the value of the projection order is increased). Moreover, since only the parameters available from the adaptive filter are required and there is no need for a priori information about the acoustic environment, the proposed algorithm is easy to control in practice. Also, it is interesting to notice that the step-size of the proposed VSS-APA does not depend explicitly on the near-end signal, even if it was developed taking into account its presence; consequently, a robust behaviour under near-end signal variations (e.g., background noise increase and double-talk) is expected.

3. SIMULATION RESULTS

The simulations were performed in an AEC context. The measured acoustic impulse response has $L = 512$ coefficients; the same length is used for the adaptive filter. The far-end signal $x(n)$ is a speech sequence. An independent white Gaussian noise signal $w(n)$ is added to the echo signal $y(n)$, with 20 dB signal-to-noise ratio (SNR). The performance is evaluated in terms of the normalized misalignment (in dB), defined as $20\log_{10}(\|\mathbf{h} - \hat{\mathbf{h}}(n)\| / \|\mathbf{h}\|)$, where $\|\bullet\|$ denotes the l_2 norm. Besides the classical APA, another two members of the APA family are chosen for comparisons, i.e., the variable regularized APA (VR-APA) recently proposed in [6] and the robust proportionate APA (R-PAPA) [9].

The VR-APA update is

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

Its variable regularization factor is given by

$$\delta(n) = \min\left\{\frac{L\sigma_x^2}{\zeta}, \frac{p\sigma_w^2\sigma_x^2 L}{\hat{\sigma}_{\mathbf{e}}^2(n) - p\sigma_w^2}\right\},$$

where ζ is a positive design parameter. The power of the background noise, σ_w^2, and the power of the input signal, σ_x^2, have to be known within the algorithm, while the term $\hat{\sigma}_{\mathbf{e}}^2(n)$ is evaluated as

$$\hat{\sigma}_{\mathbf{e}}^2(n) = \lambda\hat{\sigma}_{\mathbf{e}}^2(n-1) + (1-\lambda)\left\|\mathbf{e}(n)\right\|^2,$$

where λ is a weighting parameter as in (13). This algorithm can be also considered a VSS-APA; the experimental results presented in [6] show that it outperforms other VSS-APAs.

For the scenarios with near-end signal variations (i.e., background noise increase or double-talk), the R-PAPA is considered for comparison. This algorithm is not included in the other single-talk scenarios since the regular APA in these cases outperforms it [9]. In order to evaluate the validity of (18), an ideal version of the proposed algorithm is considered in all the simulations. This ideal VSS-APA (VSS-APA-id) is based on the assumption that the near-end signal $v(n)$ is available (which is not true in practice); its power estimate is used under the square-root from (21).

The first set of simulations is performed in a single-talk case. In Fig. 1, the proposed VSS-APA is compared with the classical APA with $\mu = 0.2$ and with the VR-APA with $\zeta = 1$, for a value of the projection order $p = 2$. The values of these parameters were found to offer a proper compromise between the convergence rate, misalignment, and complexity [10]. It is assumed that the power of the background noise is known for the VR-APA. The regularization factor for APA and VSS-APA is $\delta = 50\sigma_x^2$. The weighting factor λ is computed using $K = 6$ [7]. It can be notice from Fig. 1 that the initial convergence rate is almost the same for all the algorithms but the proposed VSS-APA achieved the lowest final misalignment, which is close to the "ideal" case.

An abrupt change of the acoustic echo path is considered in Figs. 2 and 3; the acoustic impulse response was shifted to the right by 12 samples after 21 seconds from the debut of the adaptive process. In Fig. 2, the projection order for all the algorithms is $p = 2$. As expected, the proposed algorithm has a slower tracking reaction as compared to the other algorithms, since the assumption (16) is strongly biased in this situation. Nevertheless, as it can be noticed from Fig. 3, the tracking capabilities of the VSS-APA are significantly improved for larger values of the projection order, e.g., $p = 8$. The regularization factor in this case was $\delta = 200\sigma_x^2$. The reason behind this choice is that the value of the regularization factor depends on the value of the projection order of the algorithm [1], [6]; as the value of the projection order of the APA becomes larger, the condition number of the matrix $\mathbf{X}^T(n)\mathbf{X}(n)$ also grows.

An increase of the background noise is experienced in Fig. 4; in this scenario, the SNR decreases from 20 dB to 10 dB after 14 seconds from the debut of the adaptive process, for a period of 14 seconds. It is assumed that the new background noise power estimate is not available for VR-APA. The R-PAPA is also included for comparison; its parameters are set as in [9]. It can be noticed that the proposed VSS-APA is very robust against the background noise variation, while the classical APA and VR-APA are affected by this change in the acoustic environment.

A second set of simulations is performed in a double-talk situation. In Fig. 5, the involved algorithms are APA, VR-APA, VSS-APA, and VSS-APA-id, but without using any DTD. The near-end speech appears after 14 seconds from the debut of the adaptive process, for a period of 9.2 seconds. The R-PAPA was not included in this experiment; due to its nature, this algorithm is equipped with a DTD. It can be noticed that the proposed algorithm outperforms by far the APA and the VR-APA. In practice, a simple DTD can be involved

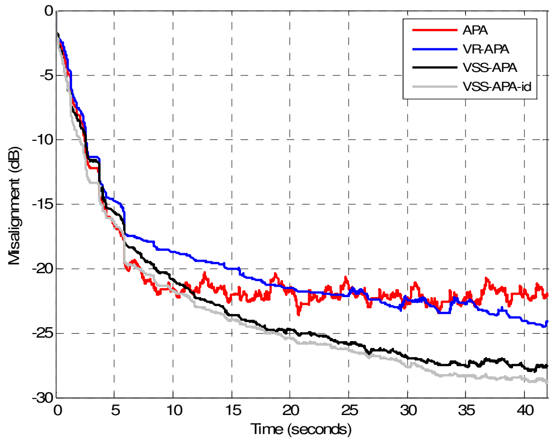

Figure 1 - Misalignments of APA with $\mu = 0.2$, VR-APA with $\zeta = 1$, VSS-APA, and VSS-APA-id. Single-talk case, $L = 512$, $p = 2$ for all the algorithms, $\delta = 50\sigma_x^2$ (for APA and VSS-APA), SNR = 20dB.

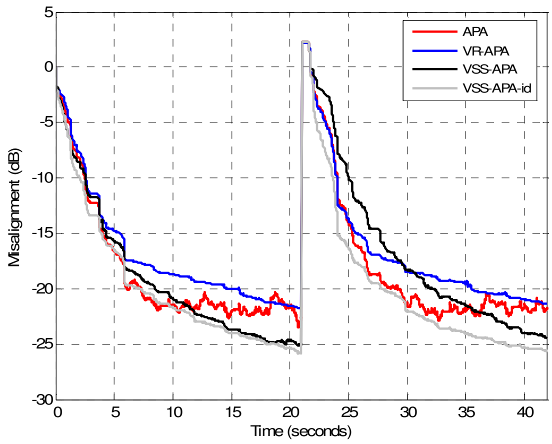

Figure 2 – Echo path changes at time 21. Other conditions are the same as in Fig. 1.

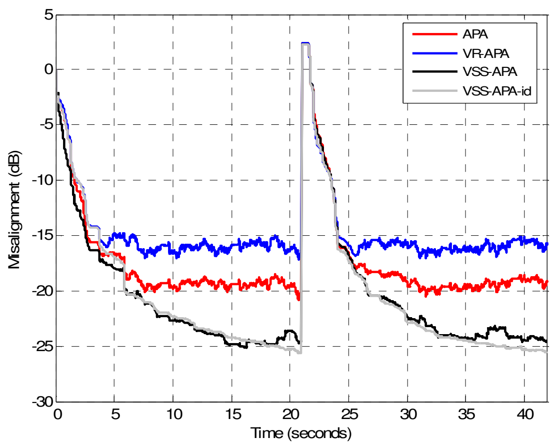

Figure 3 – Echo path changes at time 21, $p = 8$, $\delta = 200\sigma_x^2$ (for APA and VSS-APAs). Other conditions are the same as in Fig. 1.

in order to enhance the performance of the proposed algorithm during double-talk periods. In Fig. 6, the previous experiment is repeated using a Geigel DTD [11]. Its settings are chosen assuming a 6dB attenuation, i.e., the threshold is equal to 0.5 and the hangover time is set to 240 samples [9]. The algorithms chosen for comparison are APA, R-PAPA, VSS-APA, and VSS-APA-id. The VR-APA was removed from the list because it performs similarly to the APA in double-talk situations. It can be noticed from Fig. 6 that the performance of the VSS-APA is improved as compared to the previous case, and it outperforms the R-PAPA. The APA can not be "saved" by this procedure; it requires more complex DTDs, e.g., [12]. The presence of the DTD does not influence the performance of the VSS-APA-id.

4. CONCLUSIONS

The basic idea of the classical APA, i.e., to cancel p a posteriori errors, was modified within the proposed VSS-APA in order to take into account the existence of the near-end signal. The variable step-size formula of the proposed algorithm does not require any additional parameters from the acoustic environment. Consequently, the proposed VSS-APA is very suitable in practice. As compared to other APAs, it was found to be more robust to near-end signal variations like the increase of the background noise or double-talk.

REFERENCES

[1] J. Benesty, T. Gaensler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*. Berlin, Germany: Springer-Verlag, 2001.

[2] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electron. Commun. Jpn.*, vol. 67-A, no. 5, pp. 19-27, May 1984.

[3] S. L. Gay and S. Tavathia, "The fast affine projection algorithm," in *Proc. IEEE ICASSP 1995*, Detroit, U.S.A., May 1995, vol. 5, pp. 3023-3026.

[4] F. Albu and C. Kotropoulos, "Modified Gauss-Seidel affine projection algorithm for acoustic echo cancellation," in *Proc. IEEE ICASSP 2005*, Philadelphia, U.S.A., Apr. 2005, vol. 3, pp. 121-124.

[5] H.-C. Shin, A. H. Sayed, and W.-J. Song, "Variable step-size NLMS and affine projection algorithms," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 132–135, Feb. 2004.

[6] H. Rey, L. Rey Vega, S. Tressens, and J. Benesty, "Variable explicit regularization in affine projection algorithm: Robustness issues and optimal choice," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2096–2108, May 2007.

[7] J. Benesty, H. Rey, L. Rey Vega, and S. Tressens, "A nonparametric VSS NLMS algorithm," *IEEE Signal Process. Lett.*, vol. 13, no. 10, pp. 581–584, Oct. 2006.

[8] C. Paleologu, S. Ciochină, and J. Benesty, "Variable step-size NLMS algorithm for under-modeling acoustic echo cancellation," *IEEE Signal Process. Lett.*, vol. 15, pp. 5–8, 2008.

[9] T. Gänsler, S. L. Gay, M. M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp. 656–663, Nov. 2000.

[10] C. Paleologu, J. Benesty, and S. Ciochină, "A variable step-size affine projection algorithm designed for acoustic echo cancellation," submitted to *IEEE Trans. Audio, Speech, and Language Processing*.

[11] D. L. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Trans. Commun.*, vol. 26, no. 5, pp. 647–653, May 1978.

[12] J. Benesty, D.R. Morgan, and J. H. Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 2, pp. 168–172, Mar. 2000.

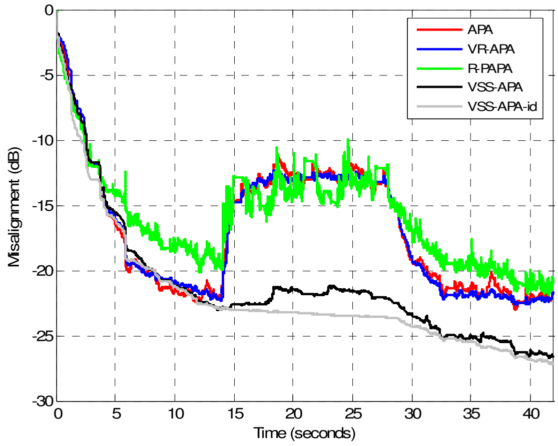

Figure 4 - Misalignments of APA, VR-APA, R-PAPA with the settings from [9], VSS-APA, and VSS-APA-id. Background noise variation at time 14, for a period of 14 seconds (SNR decreases from 20dB to 10 dB). Other conditions are the same as in Fig. 1.

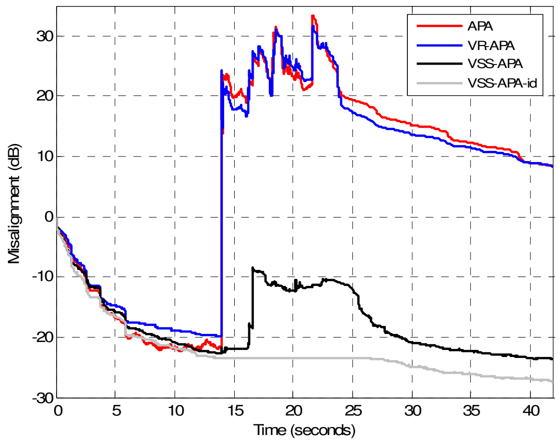

Figure 5 - Double-talk case (the near-end speech appears at time 14, for a period of 9.2 sec.), without DTD. Other conditions are the same as in Fig. 1.

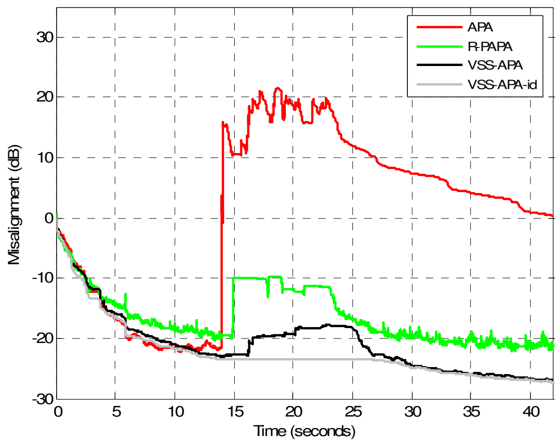

Figure 6 – Misalignments of APA, R-PAPA, VSS-APA, and VSS-APA-id. Double-talk case from Fig. 5, with Geigel DTD.