GAUSS-SEIDEL BASED VARIABLE STEP-SIZE AFFINE PROJECTION ALGORITHMS FOR ACOUSTIC ECHO CANCELLATION

Felix Albu[1], *Constantin Paleologu*[1], *and Jacob Benesty*[2]

[1] Department of Telecommunications, University Politehnica of Bucharest
1-3, Iuliu Maniu Blvd., 061071, Bucharest, Romania
email: {felix, pale}@comm.pub.ro
[2] INRS-EMT, Universite du Quebec
QC H5A 1K6, Montreal, Canada
email: benesty@emt.inrs.ca

ABSTRACT

Fast affine projection (FAP) algorithms have proved to be very attractive choice for acoustic echo cancellation (AEC). These algorithms offer a good trade-off between convergence rate and computational complexity. Most of the existing FAP algorithms use a constant step-size and need to compromise between several performance criteria (e.g., fast convergence and low misalignment). In this paper, two FAP algorithms based on the Gauss-Seidel method and using a variable step-size (VSS) are proposed for AEC. It is shown that the proposed algorithms are more robust against near-end signal variations (including double-talk) than their counterparts that use a fixed step-size.

1. INTRODUCTION

In echo cancellation systems, an adaptive filter algorithm is used to reduce the echo. The well-known normalized least-mean-square (NLMS) algorithm has been widely used in this context. Nevertheless, it converges slowly for acoustic echo cancellation (AEC) applications, where long length adaptive filters are used in order to model the acoustic echo paths. The affine projection algorithm (APA) [1] can be considered as a generalization of the NLMS algorithm that provides an improved convergence speed, especially for high-correlated signals, like speech. In terms of convergence rate, it has a performance that rivals with the more complex recursive least-squares (RLS) algorithms in many situations. However, the fast affine projection (FAP) algorithm proposed in [2] suffers from numerical instability when implemented with an embedded fast RLS algorithm. A key element in other proposed FAP algorithms is the approach to solve the encountered linear system. The choice of the approach (i.e., direct or iterative) determines the stability of the FAP algorithm.

Several proposed FAP algorithms use an approximation that leads to simpler algorithms if the step-size is $\mu = 1$ (i.e., non-relaxed case) or close to 1, e.g., $0.7 < \mu \le 1$ (see [2]–[5] and references therein). For such values, these algorithms have a fast convergence, but they exhibit a high sensitivity to noisy inputs. Two such efficient algorithms based on Gauss-Seidel iteration have been proposed in [4] (the GSFAP algorithm) and [5] [the Gauss-Seidel pseudo affine projection (GS-PAP) algorithm]. They have been applied in different application areas such as AEC [2]–[6], active noise control [7], [8], and hearing aid [9]. All these FAP versions use a fixed step-size. Better performances in adverse conditions might be expected if a variable step-size (VSS) is used.

In [10] a robust variable step-size pseudo affine projection (VSS-PAP) algorithm has been disclosed. Other variable step-size versions of the NLMS algorithm and APA have been recently proposed in [11] and [12], in the context of AEC. It was shown that they are reliable solutions in case of near-end signal variations, including double-talk. The same step-size computational method used in these VSS algorithms can be adapted to the much less computational GSFAP and GS-PAP algorithms. An alternative to the Gauss-Seidel method is to use the dichotomous coordinate descent (DCD) method proposed in [8] and [13]. However, although this DCD method reduces the number of multiplications, it increases significantly the number of additions.

The outline of the paper is as follows. The proposed VSS-GS-PAP and VSS-GSFAP algorithms are described in Section 2. In Section 3, the behaviour of these algorithms for single-talk and double-talk AEC scenarios is examined and compared with those of the algorithms using a fixed step-size; a comparison with the VSS-PAP algorithm from [10] is also made. Section 4 concludes this work.

2. VSS-GS-PAP AND VSS-GSFAP ALGORITHMS FOR AEC

The system model for an AEC scenario is depicted in Fig. 1. The far-end signal $x(n)$ goes through the echo path $\mathbf{h}$, providing the echo signal $y(n)$. This signal is added with the near-end signal $v(n)$ (which can contain both the background noise and the near-end speech), resulting the microphone signal $d(n)$. The adaptive filter, defined by the vector $\hat{\mathbf{h}}(n)$, aims to produce at its output an estimate of the echo, $\hat{y}(n)$, while the error signal $e(n)$ should contain an estimate of the near-end signal. The AEC scheme from Fig. 1 represents a "system identification" configuration, because the goal is to identify an unknown system (i.e., the acoustic echo path, $\mathbf{h}$) but it also can be viewed as an "interference cancelling" configuration, aiming to recover the near-end signal, $v(n)$, corrupted by a perturbation, i.e., the acoustic echo, $y(n)$.

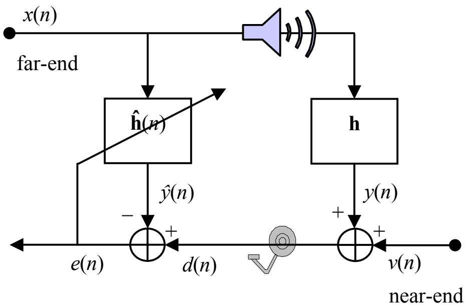

Fig. 1. AEC configuration.

Through this paper, the following notation will be used. L is the length of the adaptive filter, K is the projection order of the APAs, δ is a regularization parameter, and λ is a forgetting factor. $\mathbf{I}_K$ denotes a $K \times K$ identity matrix, $\mu(n)$ is the variable step-size at time instant n, $\mathbf{x}(n) = \left[x(n), x(n-1), ..., x(n-L+1)\right]^T$ is the input signal vector, where superscript T denotes transposition. $\mathbf{R}(n)$ is the $K \times K$ auto-correlation matrix of the input signal, $\boldsymbol{\xi}(n) = \left[x(n), x(n-1), ..., x(n-K+1)\right]^T$ is a $K \times 1$ vector, $\mathbf{u}(n) = \left[u(n), ..., u(n-L+1)\right]^T$ is the approximated vector [see (8)], $\overline{\mathbf{u}}(n)$ contains the first L – 1 elements of $\mathbf{u}(n)$ and $\mathbf{b} = \left[1 \quad \mathbf{0}^T_{(K-1)\times 1}\right]^T$, $\hat{\mathbf{h}}(n) = \left[\hat{h}_0(n), ..., \hat{h}_{L-1}(n)\right]^T$ is the adaptive filter coefficients vector, and $\mathbf{w}(n) = \left[w_0(n), ..., w_{L-1}(n)\right]^T$ is the auxiliary filter coefficient vector [2]–[4]. Also, $\mathbf{r}(n)$ is a $K \times 1$ autocorrelation vector, $\tilde{\mathbf{r}}(n)$ is the K – 1 lower part of the vector $\mathbf{r}(n)$, $\mathbf{p}(n)$ is a $K \times 1$ solution vector to the linear prediction problem ([4] – [6], [10]), and $p_i(n)$ (with i = 1, …, K) is the ith element of the vector $\mathbf{p}(n)$. In all algorithms, one Gauss-Seidel iteration is made, using $\mathbf{p}(n-1)$ as an initial approximation in order to compute $\mathbf{p}(n)$. Finally, $\mathbf{e}(n)$ is a $K \times 1$ vector, $e_K(n)$ is the last element of the vector $\mathbf{e}(n)$, and $\overline{\mathbf{e}}(n)$ is a vector consisting of the uppermost K – 1 elements of the vector $\mathbf{e}(n)$. The matrix $\mathbf{R}(n)$ is updated by replacing the first row and column with the elements of $\mathbf{r}(n)$, while the bottom-right $(K-1)\times(K-1)$ sub-matrix is replaced with the top-left $(K-1)\times(K-1)$ sub-matrix of $\mathbf{R}(n-1)$ [8], [13].

For both proposed VSS-GS-PAP and VSS-GSFAP algorithms, the following three equations are used in order to compute the variable step-size [11], [12]:

$$\hat{\sigma}_s^2(n) = \lambda \hat{\sigma}_s^2(n-1) + (1-\lambda)\left[d^2(n) - \hat{y}^2(n)\right] \quad (1)$$

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

$$\mu(n) = \left|1 - \sqrt{\frac{\hat{\sigma}_s^2(n)}{\varepsilon + \hat{\sigma}_e^2(n)}}\right| \quad (3)$$

where $\hat{\sigma}_e^2(0) = 0$, $\hat{\sigma}_s^2(0) = 0$, ε is a small positive constant, and the forgetting factor λ is evaluated as in [14]. The AEC scheme can be viewed as an "interference cancelling" configuration, aiming to recover a "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 [12]. The main feature provided by the variable step-size of the proposed algorithms is the robustness against near-end signal variations (e.g., double-talk) [12]. In the context of echo cancellation, a reasonable objective is to recover the near-end signal from the error signal of the adaptive filter. It should be noted that the classical APAs aim to cancel K previous *a posteriori* errors at every step of the algorithm; thus, the existence of the near-end signal is not taken into account. The previous relations [i.e., (1)–(3)] result by following the approach proposed in [12] and taking into account that the projection order is relatively small in AEC applications (e.g., K = 2 or 4).

The equations that define the proposed VSS-GS-PAP algorithm are:

Initialization

$$\mathbf{x}(0) = \mathbf{0}_{L\times 1},\ \mathbf{R}(0) = \delta \mathbf{I}_K,\ \mathbf{u}(0) = \mathbf{0}_{L\times 1},\ \hat{\mathbf{h}}(0) = \mathbf{0}_{L\times 1},$$
$$\mathbf{p}(0) = \mathbf{0}_{K\times 1},\ \boldsymbol{\xi}(0) = \mathbf{0}_{K\times 1},\ \mathbf{r}(0) = \left[\delta, \mathbf{0}^T_{(K-1)\times 1}\right]^T \quad (4)$$

For time index n = 1, 2, …

$$\mathbf{r}(n) = \mathbf{r}(n-1) + x(n)\boldsymbol{\xi}(n) - x(n-L)\boldsymbol{\xi}(n-L) \quad (5)$$

update $\mathbf{R}(n)$ using $\mathbf{r}(n)$ (6)

solve $\mathbf{R}(n)\mathbf{p}(n) = \mathbf{b}$ (using one Gauss-Seidel iteration) (7)

$$\mathbf{u}(n) = \left[\boldsymbol{\xi}^T(n)\mathbf{p}(n)/p_1(n) \quad \overline{\mathbf{u}}^T(n-1)\right]^T \quad (8)$$

$$\hat{y}(n) = \mathbf{x}^T(n)\hat{\mathbf{h}}(n-1) \quad (9)$$

$$e(n) = d(n) - \hat{y}(n) \quad (10)$$

compute $\mu(n)$ according to (1)–(3)

$$\overline{e}(n) = \frac{\mu(n)e(n)}{\delta + \mathbf{u}^T(n)\mathbf{x}(n)} \quad (11)$$

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mathbf{u}(n)\overline{e}(n) \quad (12)$$

The VSS-GS-PAP algorithm has $2L + K^2 + 3K + 10$ multiplications, 4 divisions, and 1 square-root operation.

More details about the GS-PAP algorithm can be found in [5].

The GSFAP algorithm [4] has several different equations than GS-PAP. Accordingly, the proposed VSS-GSFAP is defined as follows:

Initialization

$$\mathbf{x}(0)=\mathbf{0}_{L\times1},\ \mathbf{R}(0)=\delta\mathbf{I}_K,\ \mathbf{e}(0)=\mathbf{0}_{K\times1},\ \mathbf{w}(0)=\mathbf{0}_{L\times1},$$
$$\mathbf{p}(0)=\mathbf{0}_{K\times1},\ \boldsymbol{\xi}(0)=\mathbf{0}_{K\times1},\ \mathbf{r}(0)=\left[\delta,\mathbf{0}^T_{(K-1)\times1}\right]^T,\ \mu(0)=1 \quad (13)$$

For time index n = 1, 2, …

$$\mathbf{r}(n)=\mathbf{r}(n-1)+x(n)\boldsymbol{\xi}(n)-x(n-L)\boldsymbol{\xi}(n-L)$$

update $\mathbf{R}(n)$ using $\mathbf{r}(n)$

solve $\mathbf{R}(n)\mathbf{p}(n)=\mathbf{b}$ (using one Gauss-Seidel iteration)

$$\hat{y}(n)=\mathbf{x}^T(n)\mathbf{w}(n-1)+\mu(n-1)\tilde{\mathbf{r}}(n)\bar{\mathbf{e}}(n-1) \quad (14)$$
$$e(n)=d(n)-\hat{y}(n)$$
$$\mathbf{e}(n)=\left[0\quad \bar{\mathbf{e}}^T(n-1)\right]^T+e(n)\mathbf{p}(n) \quad (15)$$

compute $\mu(n)$ according to (1)–(3)

$$\mathbf{w}(n)=\mathbf{w}(n-1)+\mu(n)\mathbf{x}(n-K+1)e_K(n) \quad (16)$$

The VSS-GSFAP algorithm has $2L+K^2+4K+10$ multiplications, 2 divisions, and 1 square-root operation. More details about the GS-PAP algorithm can be found in [4].

The relations for computing the variable step-size [i.e., (1)–(3)] add 9 multiplications, 6 additions, 1 square-root, and 1 division. This additional complexity is very small in comparison with the overall complexity of both GSFAP and GS-PAP algorithms.

The VSS-PAP proposed in [10] uses a different strategy for evaluating the variable step-size parameter. It is based on the method developed in [15], i.e., minimization of the mean-square deviation (MSD). The total computational complexity of the VSS-PAP is $2L+K^2+5K+12$ multiplications and divisions.

It can be noticed that the VSS-PAP, VSS-GS-PAP, and VSS-GSFAP algorithms have roughly the same numerical complexity. As shown in [4], [5], and [9], all these considered algorithms have a much lower complexity than the original APA.

3. SIMULATIONS

The simulations were performed in an AEC context as shown in Fig. 1. The GSFAP, GS-PAP, VSS-GSFAP, VSS-PAP, and VSS-GS-PAP algorithms were compared. For the VSS-PAP algorithm the same parameters from [10] were used. The length of the adaptive filter is set to 512 coefficients. The measured car impulse response of the acoustic echo path is plotted in Fig. 2a (the sampling rate is 8 kHz); its entire length has 1024 coefficients. This length is truncated to the first 512 coefficients for most of the experiments performed in an exact modelling case.

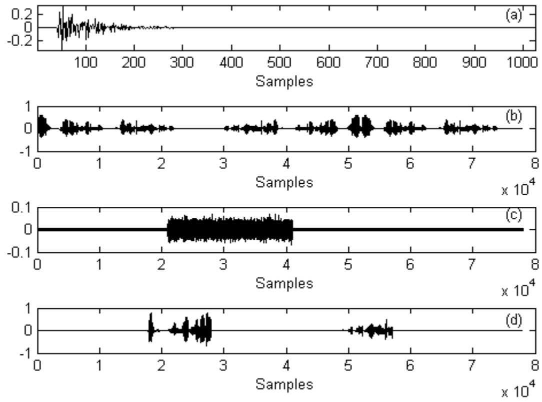

Fig. 2. (a) Measured room acoustic impulse response; (b) far-end speech signal used in the experiments; (c) the used background noise with variable SNR; (d) near-end speech signal used in the double-talk case.

Also, the entire length of the acoustic impulse response is used for one experiment performed in the under-modelling case. Its last 512 coefficients amplitude is smaller than those of the first 512 coefficients. An independent white Gaussian noise signal is added to the echo signal $y(n)$, with 30 dB signal-to-noise ratio (SNR) for most of the experiments. The value of the projection order for all simulations is $K=4$. In all the following experiments, in order to approach the context of typical AEC applications, the speech sequence from Fig. 2b is used as the far-end signal. Both single-talk and double-talk scenarios are considered. The performance for the exact modelling scenario is evaluated in terms of the normalized misalignment (in dB), defined as $20\log_{10}\left[\left\|\mathbf{h}-\hat{\mathbf{h}}(n)\right\|/\left\|\mathbf{h}\right\|\right]$, where $\|\bullet\|$ denotes the l_2 norm. In the under-modelling scenario, the expression of the normalized misalignment is evaluated by padding the vector of the adaptive filter coefficients with 512 zeros. The forgetting factor is $\lambda=1-1/(6L)$ [14]. It is well known that the GSFAP and VSS-GSFAP algorithms do not compute explicitly the adaptive filter coefficients; they compute the auxiliary filter coefficients, $\mathbf{w}(n)$. However, the true filter coefficients, $\hat{\mathbf{h}}(n)$, are computed from the auxiliary coefficients (see [2] and [3]) in order to investigate the convergence properties of the GSFAP and VSS-GSFAP algorithms.

3.1 Single-Talk Scenario

Fig. 3 shows the misalignment curves in case of a single-talk scenario. In terms of the final misalignment, it can be seen that the VSS scheme improves the overall performance of the fixed step-size Gauss-Seidel based algorithms. Also, it can be noticed that the VSS-PAP [10] is slightly outperformed by the proposed VSS algorithms.

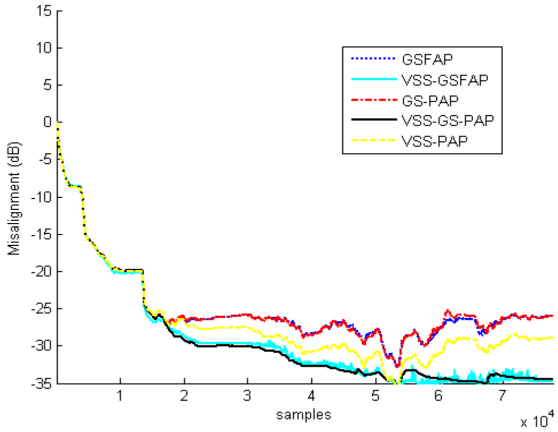

Fig. 3. Misalignment of GSFAP, GS-PAP, VSS-GSFAP, VSS-PAP, and VSS-GS-PAP algorithms. Single-talk case, $L = 512$, $K = 4$, and SNR = 30dB.

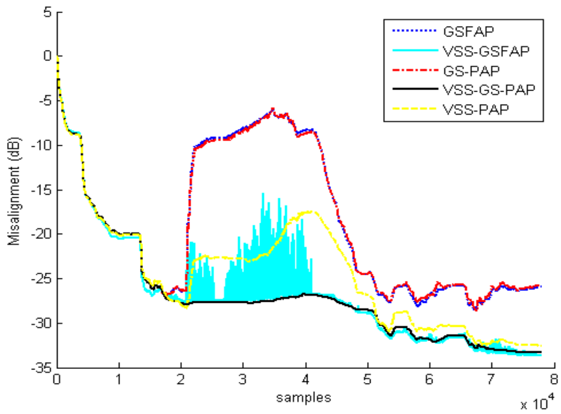

Fig. 4. Misalignments of GSFAP, GS-PAP, VSS-GSFAP, VSS-PAP, and VSS-GS-PAP algorithms. Single-talk case, variable background noise (SNR decreases from 30 dB to 10 dB for 20000 samples; see Fig. 2c), L=512, and $K = 4$.

To make more realistic results, for the second experiment we considered a variation of the background noise, i.e., the SNR decreases from 30 dB to 10 dB after 21000 samples from the debut of the adaptive process, for a period of 20000 samples (see Fig. 2c). The results are presented in Fig. 4. The behavior of GS-PAP, GSFAP, VSS-PAP, VSS-GSFAP, and VSS-GS-PAP algorithms are evaluated in the exact modeling case. It can be seen that the VSS scheme increase the robustness of both proposed algorithms to a variation of the background noise**.** The VSS-GS-PAP is superior to all the other algorithms in terms of robustness against this type of background noise variation.

Fig. 5 considers the under-modeling case (which is also a realistic case in AEC, due to the excessive length of the acoustic impulse response) and a sudden change of the acoustic path. The acoustic impulse response of Fig. 2a was shifted to the right by 12 samples after 39000 samples from the debut of the adaptive process.

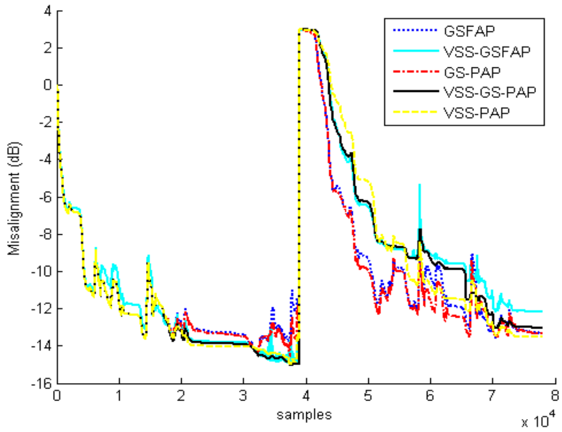

Fig.5. Misalignments of GSFAP, GS-PAP, VSS-GSFAP, VSS-PAP and VSS-GS-PAP algorithms. Single-talk case, under-modeling scenario (the acoustic impulse response has 1024 coefficients), echo path change after 39000 samples, L=512, $K = 4$, and SNR=30 dB.

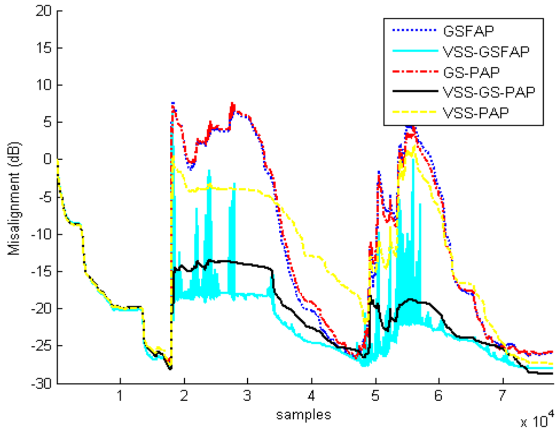

Fig. 6. Misalignments of GSFAP, GS-PAP, VSS-GSFAP, VSS-PAP and VSS-GS-PAP algorithms. Double-talk case without using DTD, L=512, K = 4, and SNR=30 dB.

As expected, it can be noticed that the re-convergence performance of the algorithms with fixed step-size (i.e., GSFAP and GS-PAP) is slightly better.

3.2 Double-Talk Scenario

Perhaps the most challenging problem in echo cancellation is the double-talk situation. Such a scenario is considered in the simulations using the speech signals from Fig. 2b and Fig. 2d. It can be noticed from Fig. 6 that the VSS-GS-PAP algorithm is superior to the VSS-PAP algorithm. The VSS-GSFAP algorithm shows a reduced robustness in this situation if compared with the VSS-GS-PAP algorithm, although its performances are superior to those of the fixed step-size algorithms. It should be noted that no double-talk detector (DTD) was used. It is expected that a DTD could improve the overall performances. The previous experiment is repeated using a simple Geigel DTD [16]. Its settings are chosen assuming a 6 dB attenuation, i.e., the threshold is equal to 0.5 and the hangover time is set to 240 samples.

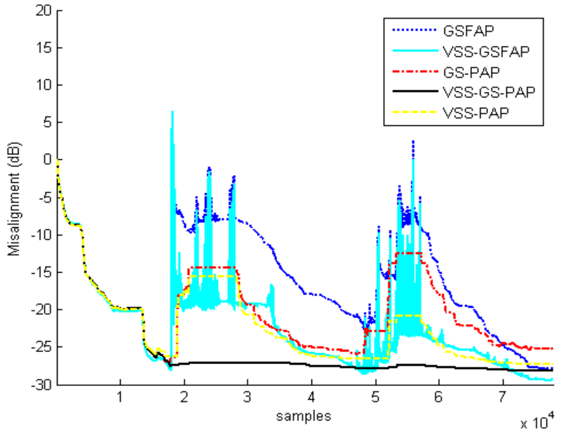

Fig. 7. Misalignments of GSFAP, GS-PAP, VSS-GSFAP, VSS-PAP, and VSS-GS-PAP algorithms. Double-talk case using Geigel DTD, L=512, K = 4, and SNR=30 dB.

Fig. 7 confirms that the algorithms performance benefit from its use. Most important, it also shows the superior robustness of the VSS-GS-PAP algorithm to the other considered algorithms. The VSS-GSFAP algorithm suffers from a certain degree of fluctuations in its convergence In several investigated situations. They are caused by the propagation of the Gauss-Seidel solution errors through the algorithm when the error signal and the computed step-size have small values.

4. CONCLUSIONS

The VSS-GS-PAP and VSS-GSFAP algorithms have been proposed for AEC. A variable step-size was used in order to take into account the existence and the non-stationarity nature of the near-end signal as well as the under-modelling noise. The simulation results performed in an AEC context showed the superior robustness to near-end signal variations (like the increase of the background noise or double-talk) of the VSS-GS-PAP algorithm in comparison with VSS-PAP and VSS-GSFAP algorithms.

Acknowledgment: This work was supported by the Romanian UEFISCSU under Grant PN-II no. 331/ 01.10.2007

REFERENCES

[1] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electronics and Communications in Japan*, vol. 67-A, no. 5, pp. 19–27, May 1984.

[2] S.L. Gay "A fast converging, low complexity adaptive filtering algorithm," in *Proc. IWAENC*, 1993, pp. 223–226.

[3] S. Oh, D. Linebarger, B. Priest, and B. Raghothaman, "A fast affine projection algorithm for an acoustic echo cancellation using a fixed-point DSP processor," in *Proc. IEEE ICASSP*, vol. 5, 1997, pp. 4121–4124.

[4] F. Albu, J. Kadlec, N. Coleman, and A. Fagan, "The Gauss-Seidel fast affine projection algorithm," in *Proc. IEEE SIPS*, 2002, pp. 109–114.

[5] F. Albu and H.K. Kwan, "Combined echo and noise cancellation based on Gauss-Seidel pseudo affine projection algorithm," in *Proc. IEEE ISCAS*, 2004, pp. 505–508.

[6] P Sheikhzadeh, H. Whyte, and K.R.L. Brennan, "Partial update subband implementation of complex pseudo-affine projection algorithm on oversampled filterbanks," in *Proc. IEEE ICASSP*, vol. IV, 2004, pp. 373–376.

[7] J.M. Wesselink and A.P. Berkhoff, "Fast affine projections and the regularized modified filtered-error algorithm in in multichannel active noise control," *Journal of the Acoustical Society of America*, vol. 124, pp. 949–960, Aug. 2008.

[8] F. Albu, M. Bouchard, and Y. Zakharov, "Pseudo-affine projection algorithms for multichannel active noise control," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 15 no. 3, pp. 1044–1052, Mar. 2007.

[9] S. Lee, I. Kim, and Y. Park, "Approximated affine projection algorithm for feedback cancellation in hearing aids," *Computer Methods and Programs in Biomedicine*, vol. 87, pp. 254–261, Sept. 2007.

[10] L. Lee, Y.-C. Park, D.-H. Youn, "Robust pseudo affine projection algorithm with variable step-size," *Electronics Letters*, vol. 44, pp. 250–252, Jan. 2008.

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

[12] C. Paleologu, J. Benesty, and S. Ciochină, "A Variable step-size affine projection algorithm designed for acoustic echo cancellation," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 16, pp. 597–600, Nov. 2008.

[13] Y. Zakharov and F. Albu, "Coordinate descent iterations in fast affine projection algorithm," *IEEE Signal Processing Letters*, vol. 12, pp. 353–356, May 2005.

[14] J. Benesty, H. Rey, L. Rey Vega, and S. Tressens, "A nonparametric VSS NLMS algorithm," *IEEE Signal Processing Letters*, vol. 13, pp. 581–584, Oct. 2006.

[15] H.-C. Shin, A. H. Sayed, and W.-J. Song, "Variable step-size NLMS and affine projection algorithms," *IEEE Signal Processing Letters*, vol. 11, pp. 132–135, Feb. 2004.

[16] D. L. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Trans. on Communications*, vol. 26, pp. 647–653, May 1978.