VARIABLE-FORGETTING FACTOR RLS FOR STEREOPHONIC ACOUSTIC ECHO CANCELLATION WITH WIDELY LINEAR MODEL

Cristian Stanciu[1], *Constantin Paleologu*[1], *Jacob Benesty*[2], *Silviu Ciochina*[1], *and Felix Albu*[3]

[1] University Politehnica of Bucharest, Romania, e-mail: {cristian, pale, silviu}@comm.pub.ro
[2] INRS-EMT, University of Quebec, Montreal, Canada, e-mail: benesty@emt.inrs.ca
[3] Valahia University of Targoviste, Romania, e-mail: felix.albu@valahia.ro

ABSTRACT

A widely linear (WL) model was recently proposed for stereophonic acoustic echo cancellation (SAEC). In this framework, the classical two-input/two-output SAEC scheme with real random variables is recasted as a single-input/single-output system with complex random variables. The main advantage of this approach is that instead of handling two (real) output signals separately, we only handle one (complex) output signal. In general, due to their good convergence features, recursive least-squares (RLS) algorithms are preferable for SAEC applications. However, the performance of RLS-based algorithms is governed by the forgetting factor, whose value leads to a compromise between convergence rate/tracking capabilities on the one hand and misadjustment/stability on the other hand. In this paper, we develop a variable-forgetting factor RLS (VFF-RLS) algorithm for SAEC with the WL model. Simulation results indicate the good performance (in terms of tracking and robustness) of the proposed algorithm.

***Index Terms*—** Stereophonic acoustic echo cancellation (SAEC), widely linear (WL) model, recursive least-squares (RLS), variable-forgetting factor (VFF).

1. INTRODUCTION

The stereo transmission is becoming very popular in hands-free teleconferencing systems since it gives a realistic presence that actual single-channel systems cannot offer [1]. In this context, stereophonic acoustic echo cancellation (SAEC) is necessary for full-duplex quality communication [2]. In the usual approach, an SAEC system consists of four adaptive filters aiming at identifying four echo paths from two loudspeakers to two microphones. In other words, for each microphone in the receiving (i.e., near-end) location, the SAEC consists of the identification of a two-input unknown system, consisting of the parallel combination of two acoustic echo paths (from the two loudspeakers to the microphone).

Recently, a different approach for the SAEC problem was proposed [3] by using the widely linear (WL) model [4]. Basically, the classical two-input/two-output system with real random variables is recasted as a single-input/single-output system with complex random variables. Thus, the four real-valued acoustic impulse responses are converted to one complex-valued impulse response. The advantage of this approach is that instead of handling two (real) output signals separately, we only handle one (complex) output signal.

However, the main challenge of SAEC is that the loudspeaker (input) signals are linearly related through non-invertible acoustic room responses. The consequence of this linear relationship is that the underlying normal (or Wiener-Hopf) equations to be solved by the adaptive algorithm is a very ill-conditioned or, in the worst case, a singular problem [5]. The problem can be solved by manipulating the signals transmitted to the near-end room, e.g., using a preprocessor on the loudspeaker signals to reduce their coherence before the SAEC as well as before transmitting them to the far-end location (but without affecting much the stereo perception and the sound quality) [6]. Moreover, using an algorithm tailored to exploit the cross-correlation between the channels could mitigate the effects of the ill-conditioned normal equations to be solved. From this point of view, a natural choice of algorithm is the recursive least-squares (RLS) [7], [8].

The performance of the RLS algorithm is governed by the forgetting factor, whose value leads to a compromise between convergence rate/tracking capabilities on the one hand and misadjustment/stability on the other hand. In order to properly address this compromise, we present in this paper a variable-forgetting factor RLS (VFF-RLS) algorithm for SAEC with the WL model. The rest of the paper is organized as follows. Section 2 presents the WL model for SAEC and reviews the nonuniqueness problem in this context. In Section 3, we derive the RLS algorithm in the WL framework. The mechanism that controls the forgetting factor is developed in Section 4. The simulation results presented in Section 5 illustrate the performance of the proposed algorithm. Finally, the conclusions are summarized in Section 6.

This work was supported under the Grant POSDRU/107/1.5/S/76903 and Grant UEFISCDI PN-II-ID-PCE-2011-3-0097.

2. THE WIDELY LINEAR MODEL FOR SAEC

The classical stereophonic setup can be viewed as a two-input/two-output system with real random variables. In this scheme, there are two input or loudspeaker signals denoted by $x_{\rm L}(n)$ and $x_{\rm R}(n)$ (i.e., "left" and "right"), and two output or microphone signals denoted by $d_{\rm L}(n)$ and $d_{\rm R}(n)$, where n is the time index. The microphone signals in the receiving (i.e., near-end) location are obtained as

$$d_{\rm L}(n) = y_{\rm L}(n) + v_{\rm L}(n), \quad (1)$$
$$d_{\rm R}(n) = y_{\rm R}(n) + v_{\rm R}(n), \quad (2)$$

where $y_{\rm L}(n)$ and $y_{\rm R}(n)$ denote the stereo echo signals, and $v_{\rm L}(n)$ and $v_{\rm R}(n)$ are the near-end signals (i.e., noise or a combination of noise and near-end speech). In the near-end location, the echo signals can be modelled as [5], [6]

$$y_{\rm L}(n) = \mathbf{h}_{\rm t,LL}^T \mathbf{x}_{\rm L}(n) + \mathbf{h}_{\rm t,RL}^T \mathbf{x}_{\rm R}(n), \quad (3)$$
$$y_{\rm R}(n) = \mathbf{h}_{\rm t,LR}^T \mathbf{x}_{\rm L}(n) + \mathbf{h}_{\rm t,RR}^T \mathbf{x}_{\rm R}(n), \quad (4)$$

where $\mathbf{h}_{\rm t,LL}, \mathbf{h}_{\rm t,RL}, \mathbf{h}_{\rm t,LR}, \mathbf{h}_{\rm t,RR}$ are L-dimensional vectors of the loudspeaker-to-microphone ("true") acoustic impulse responses, the superscript ^{T} denotes transposition, and

$$\mathbf{x}_{\rm L}(n) = \begin{bmatrix} x_{\rm L}(n) & x_{\rm L}(n-1) & \cdots & x_{\rm L}(n-L+1) \end{bmatrix}^T$$
$$\mathbf{x}_{\rm R}(n) = \begin{bmatrix} x_{\rm R}(n) & x_{\rm R}(n-1) & \cdots & x_{\rm R}(n-L+1) \end{bmatrix}^T$$

comprise the L most recent loudspeaker signal samples. In order to cancel the echo, we need to estimate the four acoustic impulse responses, $\mathbf{h}_{\rm t,LL}, \mathbf{h}_{\rm t,RL}, \mathbf{h}_{\rm t,LR}, \mathbf{h}_{\rm t,RR}$, from the microphone signals $d_{\rm L}(n)$ and $d_{\rm R}(n)$.

In [3], we proposed to recast the classical two-input/two-output SAEC scheme (with real random variables) as a single-input/single-output system with complex random variables (CRVs). First, we can form the CRV:

$$d(n) = d_{\rm L}(n) + j d_{\rm R}(n) = y(n) + v(n), \quad (5)$$

where $j = \sqrt{-1}$, $y(n) = y_{\rm L}(n) + j y_{\rm R}(n)$, and $v(n) = v_{\rm L}(n) + j v_{\rm R}(n)$. Let us define the complex random vector:

$$\mathbf{x}(n) = \mathbf{x}_{\rm L}(n) + j \mathbf{x}_{\rm R}(n). \quad (6)$$

Therefore, the (complex) echo signal can be obtained as

$$y(n) = \mathbf{h}_{\rm t}^H \mathbf{x}(n) + \mathbf{h}_{\rm t}'^H \mathbf{x}^*(n), \quad (7)$$

where the superscripts ^{H} and * denote transpose-conjugate and conjugate, respectively, and

$$\mathbf{h}_{\rm t} = \mathbf{h}_{\rm t,1} + j \mathbf{h}_{\rm t,2}, \quad (8)$$
$$\mathbf{h}_{\rm t}' = \mathbf{h}_{\rm t,1}' + j \mathbf{h}_{\rm t,2}', \quad (9)$$

with $\mathbf{h}_{\rm t,1} = (\mathbf{h}_{\rm t,LL} + \mathbf{h}_{\rm t,RR})/2$, $\mathbf{h}_{\rm t,2} = (\mathbf{h}_{\rm t,RL} - \mathbf{h}_{\rm t,LR})/2$, $\mathbf{h}_{\rm t,1}' = (\mathbf{h}_{\rm t,LL} - \mathbf{h}_{\rm t,RR})/2$, and $\mathbf{h}_{\rm t,2}' = -(\mathbf{h}_{\rm t,RL} + \mathbf{h}_{\rm t,LR})/2$. In this framework, we can express (7) as

$$y(n) = \widetilde{\mathbf{h}}_{\rm t}^H \widetilde{\mathbf{x}}(n), \quad (10)$$

where $\widetilde{\mathbf{h}}_{\rm t} = \begin{bmatrix} \mathbf{h}_{\rm t}^T \ \mathbf{h}_{\rm t}'^T \end{bmatrix}^T$ and $\widetilde{\mathbf{x}}(n) = \begin{bmatrix} \mathbf{x}^T(n) \ \mathbf{x}^H(n) \end{bmatrix}^T$. Consequently, the complex observation is

$$d(n) = \widetilde{\mathbf{h}}_{\rm t}^H \widetilde{\mathbf{x}}(n) + v(n). \quad (11)$$

It can be noticed that we are dealing now with a complex acoustic impulse response of length $2L$, i.e., $\widetilde{\mathbf{h}}_{\rm t}$, whose complex input and output are, respectively, $x(n)$ and $d(n)$. From (7) or (10), we recognize the widely linear (WL) model for CRVs proposed in [4]. Thanks to the WL model, the two-input/two-output system with real random variables was converted to a single-input/single-output system with CRVs. This approach is in line with the duality principle [9].

As discussed in the previous section, it may be required to distort the input signals $x_{\rm L}(n)$ and $x_{\rm R}(n)$, in order to reduce the coherence between these two signals. However, this distortion must be performed in such a way that the quality of the signals and the stereo effect are not degraded. A simple but efficient method uses positive and negative half-wave rectifiers on each channel respectively [6], i.e.,

$$x_{\rm L}'(n) = x_{\rm L}(n) + 0.5\alpha_{\rm r} \left[x_{\rm L}(n) + |x_{\rm L}(n)| \right], \quad (12)$$
$$x_{\rm R}'(n) = x_{\rm R}(n) + 0.5\alpha_{\rm r} \left[x_{\rm R}(n) - |x_{\rm R}(n)| \right], \quad (13)$$

where $\alpha_{\rm r}$ is a parameter used to control the amount of nonlinearity. In the framework of the WL model, the complex input signal can be expressed as

$$x(n) = x_{\rm L}(n) + j x_{\rm R}(n) = e^{j\theta_{\rm r}(n)} |x(n)|, \quad (14)$$

where $\theta_{\rm r}(n)$ [with $\tan \theta_{\rm r}(n) = x_{\rm R}(n)/x_{\rm L}(n)$] and $|x(n)| = \sqrt{x_{\rm L}^2(n) + x_{\rm R}^2(n)}$ are the phase and module of $x(n)$, respectively. In this formulation, we represent the stereo perception with $\theta_{\rm r}(n)$ and the quality of the stereo signals with $|x(n)|$. A modification of $\theta_{\rm r}(n)$ only, will mostly affect the stereo effect of $x(n)$, while a modification of $|x(n)|$ will mostly affect the quality of the stereo signals.

Using the complex notation, (12) and (13) can be expressed as

$$x'(n) = x_{\rm L}'(n) + j x_{\rm R}'(n) = e^{j\theta_{\rm r}'(n)} |x'(n)|, \quad (15)$$

where $\tan \theta_{\rm r}'(n) = x_{\rm R}'(n)/x_{\rm L}'(n)$ and $|x'(n)| = \sqrt{x_{\rm L}'^2(n) + x_{\rm R}'^2(n)}$. In order to preserve the quality of the stereo signals, we should not modify the module of the complex input signal $x(n)$, but only its phase. Consequently, the following transformations can be used with the WL model [3]:

$$x_{\rm L}''(n) = \cos \theta_{\rm r}'(n) |x(n)|, \quad (16)$$
$$x_{\rm R}''(n) = \sin \theta_{\rm r}'(n) |x(n)|, \quad (17)$$

where the phase $\theta_{\rm r}'(n)$ is computed from the half-wave rectifiers [see (15)] while the module corresponds to the module of the original signals.

3. RLS ALGORITHM IN THE WL CONTEXT

In order to cancel the echo, we need to estimate the system $\widetilde{\mathbf{h}}_{\rm t}$. Let $\widetilde{\mathbf{h}}(n)$ be an adaptive filter and let

$$e(n) = d(n) - \widetilde{\mathbf{h}}^H(n-1)\widetilde{\mathbf{x}}(n) = d(n) - \hat{y}(n) \quad (18)$$

be the error signal at time n. In the following, we slightly change the notation for convenience. Let us redefine the input signal vector (of length $2L$) as

$$\widetilde{\mathbf{x}}(n) = \left[\begin{array}{ccc} \boldsymbol{\chi}^T(n) & \boldsymbol{\chi}^T(n-1) & \cdots \boldsymbol{\chi}^T(n-L+1) \end{array} \right]^T, \quad (19)$$

where $\boldsymbol{\chi}(n) = \left[\begin{array}{cc} x(n) & x^*(n) \end{array} \right]^T$. Thus, the new definitions of the true impulse response and the adaptive filter are

$$\begin{aligned} \widetilde{\mathbf{h}}_{\rm t} &= \left[\begin{array}{ccccc} h_{{\rm t},0} & h'_{{\rm t},0} & \cdots & h_{{\rm t},L-1} & h'_{{\rm t},L-1} \end{array} \right]^T, \\ \widetilde{\mathbf{h}}(n) &= \left[\begin{array}{ccccc} h_0(n) & h'_0(n) & \cdots & h_{L-1}(n) & h'_{L-1}(n) \end{array} \right]^T, \end{aligned}$$

where $h_{{\rm t},l}$, $h'_{{\rm t},l}$, $h_l(n)$, and $h'_l(n)$, with $l = 0, 1, \ldots, L-1$, are the elements of the vectors $\mathbf{h}_{\rm t}$, $\mathbf{h}'_{\rm t}$, $\mathbf{h}(n)$, and $\mathbf{h}'(n)$, respectively (i.e., the vectors are interleaved now instead of being concatenated). However, these new definitions do not change the definitions of the complex observation $d(n)$ and the error signal $e(n)$, given in (11) and (18), respectively.

In this context, the least-squares (LS) error criterion can be defined as [7]

$$J\left[\widetilde{\mathbf{h}}(n)\right] = \sum_{i=1}^{n} \lambda^{n-i} \left| d(i) - \widetilde{\mathbf{h}}^H(n)\widetilde{\mathbf{x}}(i) \right|^2, \quad (20)$$

where λ ($0 \ll \lambda < 1$) is the forgetting factor, which influences the memory of the data in the different statistics estimates. Next, we can express (20) as

$$\begin{aligned} J\left[\widetilde{\mathbf{h}}(n)\right] &= \sum_{i=1}^{n} \lambda^{n-i} |d(i)|^2 - \widetilde{\mathbf{h}}^H(n)\mathbf{p}_{\widetilde{\mathbf{x}}d}(n) \quad (21) \\ &\quad - \mathbf{p}_{\widetilde{\mathbf{x}}d}^H(n)\widetilde{\mathbf{h}}(n) + \widetilde{\mathbf{h}}^H(n)\mathbf{R}_{\widetilde{\mathbf{x}}}(n)\widetilde{\mathbf{h}}(n), \end{aligned}$$

where $\mathbf{R}_{\widetilde{\mathbf{x}}}(n) = \sum_{i=1}^{n} \lambda^{n-i}\widetilde{\mathbf{x}}(i)\widetilde{\mathbf{x}}^H(i)$ and $\mathbf{p}_{\widetilde{\mathbf{x}}d}(n) = \sum_{i=1}^{n} \lambda^{n-i}\widetilde{\mathbf{x}}(i)d^*(i)$. The minimization of $J\left[\widetilde{\mathbf{h}}(n)\right]$ with respect to $\widetilde{\mathbf{h}}(n)$ leads to the normal equations [7]:

$$\mathbf{R}_{\widetilde{\mathbf{x}}}(n)\widetilde{\mathbf{h}}(n) = \mathbf{p}_{\widetilde{\mathbf{x}}d}(n). \quad (22)$$

Assuming that $\mathbf{R}_{\widetilde{\mathbf{x}}}(n) > 0$ (i.e., the matrix is positive definite), the optimal filter in the LS sense can be deduced as $\widetilde{\mathbf{h}}(n) = \mathbf{R}_{\widetilde{\mathbf{x}}}^{-1}(n)\mathbf{p}_{\widetilde{\mathbf{x}}d}(n)$. Also, we can compute recursively:

$$\begin{aligned} \mathbf{p}_{\widetilde{\mathbf{x}}d}(n) &= \lambda \mathbf{p}_{\widetilde{\mathbf{x}}d}(n-1) + \widetilde{\mathbf{x}}(n)d^*(n), \quad (23) \\ \mathbf{R}_{\widetilde{\mathbf{x}}}(n) &= \lambda \mathbf{R}_{\widetilde{\mathbf{x}}}(n-1) + \widetilde{\mathbf{x}}(n)\widetilde{\mathbf{x}}^H(n). \quad (24) \end{aligned}$$

By applying the Woodbury's identity in (24), the inverse of $\mathbf{R}_{\widetilde{\mathbf{x}}}(n)$ can be expressed as

$$\begin{aligned} &\mathbf{R}_{\widetilde{\mathbf{x}}}^{-1}(n) \\ &= \lambda^{-1}\mathbf{R}_{\widetilde{\mathbf{x}}}^{-1}(n-1) - \lambda^{-1}\widetilde{\mathbf{k}}(n)\widetilde{\mathbf{x}}^H(n)\mathbf{R}_{\widetilde{\mathbf{x}}}^{-1}(n-1), \end{aligned} \quad (25)$$

where

$$\widetilde{\mathbf{k}}(n) = \frac{\mathbf{R}_{\widetilde{\mathbf{x}}}^{-1}(n-1)\widetilde{\mathbf{x}}(n)}{\lambda + \widetilde{\mathbf{x}}^H(n)\mathbf{R}_{\widetilde{\mathbf{x}}}^{-1}(n-1)\widetilde{\mathbf{x}}(n)} \quad (26)$$

is the Kalman gain vector. Rearranging (26) and taking (25) into account, the Kalman gain vector can be also expressed as

$$\widetilde{\mathbf{k}}(n) = \mathbf{R}_{\widetilde{\mathbf{x}}}^{-1}(n)\widetilde{\mathbf{x}}(n). \quad (27)$$

Consequently, the solution of (22) can be rewritten as

$$\widetilde{\mathbf{h}}(n) = \lambda \mathbf{R}_{\widetilde{\mathbf{x}}}^{-1}(n)\mathbf{p}_{\widetilde{\mathbf{x}}d}(n-1) + \mathbf{R}_{\widetilde{\mathbf{x}}}^{-1}(n)\widetilde{\mathbf{x}}(n)d^*(n). \quad (28)$$

Substituting (25) into the first term only on the right-hand side of (28) and using (18), the RLS filter update results as

$$\widetilde{\mathbf{h}}(n) = \widetilde{\mathbf{h}}(n-1) + \widetilde{\mathbf{k}}(n)e^*(n). \quad (29)$$

Finally, we could notice that this algorithm can be seen as a two-channel RLS, since $x(n)$ and $x^*(n)$ correspond to the inputs of these two channels.

4. VFF-RLS ALGORITHM

The performance of most RLS-type algorithms in terms of convergence rate, tracking, misadjustment, and stability depends on the forgetting factor λ. It is known that when the forgetting factor is very close to one, the algorithm achieves low misadjustment and good stability, but its tracking capabilities are reduced. A smaller value of λ improves the tracking but increases the misadjustment, and it could affect the stability of the algorithm. This was the motivation behind the development of the variable forgetting factor RLS (VFF-RLS) algorithms, e.g., [10], [11].

The RLS algorithm presented in the previous section uses a constant value of λ and needs to compromise between the previous performance criteria. In order to develop a VFF version of this algorithm, let us remember that the WL SAEC application is basically a system identification problem. The complex reference signal $d(n)$ given in (11) contains the complex echo signal $y(n)$ [see (10)] and the CRV $v(n)$ corresponding to the near-end signal. In this context, the goal of the adaptive filter is not to annihilate the error signal, but to recover the near-end signal from the error signal of the adaptive filter after this one converges to the true solution [12].

As we can notice, the signal $e(n)$ from (18) is the a priori error, since it is computed using the adaptive filter at time $n-1$. In the context of the WL model, the a posteriori error signal (based on the adaptive filter at time n) is defined as

$$\epsilon(n) = d(n) - \widetilde{\mathbf{h}}^H(n)\widetilde{\mathbf{x}}(n). \quad (30)$$

Using (18) and (29) in (30), the relation between the a posteriori and a priori errors results as

$$\epsilon(n) = e(n)\left[1 - \widetilde{\mathbf{k}}^H(n)\widetilde{\mathbf{x}}(n)\right]. \quad (31)$$

Let $u(n) = \widetilde{\mathbf{x}}^H(n)\mathbf{R}_{\widetilde{\mathbf{x}}}^{-1}(n-1)\widetilde{\mathbf{x}}(n)$, which is a real variable since the matrix $\mathbf{R}_{\widetilde{\mathbf{x}}}^{-1}(n-1)$ is Hermitian. Using this notation and taking (26) into account, (31) can be rewritten as

$$\epsilon(n) = e(n)\frac{\lambda}{\lambda + u(n)}. \quad (32)$$

As explained before, it is desirable to recover the near-end signal from the error signal. This condition can be imposed in terms of power estimates as [11]

$$E\left[|\epsilon(n)|^2\right] = \sigma_v^2(n), \quad (33)$$

where $\sigma_v^2(n) = E\left[|v(n)|^2\right]$ is the variance of $v(n)$. Multiplying (32) by its conjugate, then taking the expectations, and assuming that the input and error signals are uncorrelated (which is true when the adaptive filter has started to converge to the true solution), the condition (33) becomes

$$\lambda^2\left[\sigma_e^2(n) - \sigma_v^2(n)\right] - 2\lambda\sigma_v^2(n)\sigma_u(n) - \sigma_v^2(n)\sigma_u^2(n) = 0, \quad (34)$$

where $\sigma_e^2(n) = E\left[|e(n)|^2\right]$ and $\sigma_u^2(n) = E\left[u^2(n)\right]$ are the variances of $e(n)$ and $u(n)$, respectively. At this point, let us assume that the forgetting factor is deterministic and time dependent. Consequently, by solving the quadratic equation (34), it results the variable forgetting factor

$$\lambda(n) = \frac{\sigma_u(n)\sigma_v(n)}{\sigma_e(n) - \sigma_v(n)}. \quad (35)$$

In order to evaluate (35) in practice, we need to estimate the parameters $\sigma_e^2(n)$, $\sigma_u^2(n)$, and $\sigma_v^2(n)$. Since the signals $e(n)$ and $u(n)$ are available within the algorithm, their power estimates can be recursively evaluated as

$$\hat{\sigma}_e^2(n) = \alpha\hat{\sigma}_e^2(n-1) + (1-\alpha)|e(n)|^2, \quad (36)$$
$$\hat{\sigma}_u^2(n) = \alpha\hat{\sigma}_u^2(n-1) + (1-\alpha)u^2(n), \quad (37)$$

where $\alpha = 1 - 1/(2KL)$ $(K \geq 1)$ and $\hat{\sigma}_e^2(0) = \hat{\sigma}_u^2(0) = 0$.

The estimation of $\sigma_v^2(n)$ is more challenging. The most difficult situation is the double-talk case, when the near-end signal is a combination of noise and near-end speech. In this scenario, the parameter $\sigma_v^2(n)$ could be estimated as proposed in [13]; assuming that the adaptive filter has converged to a certain degree, the near-end signal power can be evaluated as

$$\hat{\sigma}_v^2(n) = \left|\hat{\sigma}_d^2(n) - \hat{\sigma}_{\hat{y}}^2(n)\right|, \quad (38)$$

where $\hat{\sigma}_d^2(n)$ and $\hat{\sigma}_{\hat{y}}^2(n)$ are the power estimates of $d(n)$ and $\hat{y}(n)$ [see (18)], respectively, computed similarly to (36).

Having all the parameters required for computation of $\lambda(n)$, we can now get back to (35) in order to evaluate the overall behavior of the forgetting factor. Let us examine the denominator of (35), where, theoretically, $\sigma_e(n) \geq \sigma_v(n)$. However, since power estimates are used, this condition could be biased in practice; alternatively, we can take the absolute value of this denominator. Clearly, before the algorithm converges or when there is an abrupt change of the system, $\sigma_e(n)$ is large as compared to $\sigma_v(n)$. Therefore, the denominator of (35) increases so that $\lambda(n)$ goes to lower values, providing fast convergence and tracking. On the other hand, when the algorithm converges to the steady-state solution, $\sigma_e(n) \approx \sigma_v(n)$. In order to avoid any numerical problems in (35), we can limit the value of $\lambda(n)$ to an upper bound λ_{max} (very close to one). In this sense, in order to overcome extra computational load, we can also impose $\lambda(n) = \lambda_{\text{max}}$ when the algorithm is in steady-state, by checking the condition

$$\hat{\sigma}_e(n) \leq \gamma\hat{\sigma}_v(n), \quad (39)$$

where $1 < \gamma \leq 2$ [11]. Consequently, when (39) is fulfilled, the value of the forgetting factor goes to its upper bound, thus providing low misadjustment. Summarizing the previous discussion, (35) can be rewritten in practice as

$$\lambda(n) = \begin{cases} \min\left[\frac{\hat{\sigma}_u(n)\hat{\sigma}_v(n)}{\varepsilon + |\hat{\sigma}_e(n) - \hat{\sigma}_v(n)|}, \lambda_{\text{max}}\right], & \text{if } \hat{\sigma}_e(n) > \gamma\hat{\sigma}_v(n) \\ \lambda_{\text{max}}, & \text{if } \hat{\sigma}_e(n) \leq \gamma\hat{\sigma}_v(n) \end{cases} \quad (40)$$

where the small positive constant ε prevents division by zero. The additional computational amount introduced by the VFF approach is much lower as compared to the overall complexity of the RLS algorithm.

5. SIMULATION RESULTS

Simulations are performed in the context of the WL model for SAEC. The acoustic impulse responses in the far-end location have 2048 coefficients, while the length of the impulse responses in the near-end location [i.e., $h_{\text{t,LL}}(n)$, $h_{\text{t,RL}}(n)$, $h_{\text{t,LR}}(n)$, and $h_{\text{t,RR}}(n)$] is $L = 512$. The length of the adaptive filter $\widetilde{\mathbf{h}}(n)$ is $2L = 1024$ and the sampling rate is 8 kHz. The source signal in the far-end location is a female speech sequence. The microphone signals are preprocessed based on (16) and (17), using $\alpha_{\text{r}} = 0.3$.

The proposed VFF-RLS algorithm is compared with the RLS algorithm using three constant forgetting factors, i.e., $\lambda = 1 - 1/(2L)$, $\lambda = 1 - 1/(10L)$, and $\lambda_{\text{max}} = 0.99999$. The algorithms use the same initialization $\mathbf{R}^{-1}(0) = 100\mathbf{I}_{2L}$, where $\mathbf{I}_{2L}$ is the identity matrix of size $2L \times 2L$. The specific parameter of the VFF-RLS algorithm is set to $\gamma = 1.5$. Two performance measures are used: (i) the normalized misalignment (in dB), defined as $20\log_{10}\left\|\widetilde{\mathbf{h}}_{\text{t}} - \widetilde{\mathbf{h}}(n)\right\|_2 / \left\|\widetilde{\mathbf{h}}_{\text{t}}\right\|_2$ (with $\|\cdot\|_2$ denoting the ℓ_2 norm) and (ii) the mean-square error (MSE) averaged over 256 points for the purpose of smoothing the results.

The first experiment is performed in a single-talk scenario, i.e., the near-end signal $v(n)$ consists only of the background noise. The stereo echo-to-noise ratio (SENR) is defined as $\text{SENR} = \sigma_y^2/\sigma_v^2$, where $\sigma_y^2 = E[|y(n)|^2]$ is the variance of $y(n)$. In our simulations, the background noise in the

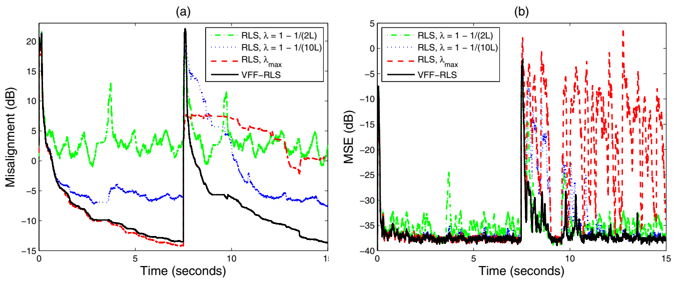

Fig. 1. Performance of the RLS algorithms using $\lambda = 1 - 1/(2L)$, $\lambda = 1 - 1/(10L)$, and $\lambda_{\text{max}} = 0.99999$, and VFF-RLS algorithm; (a) Misalignment and (b) MSE. Single-talk scenario, echo path changes at time 7.5 seconds.

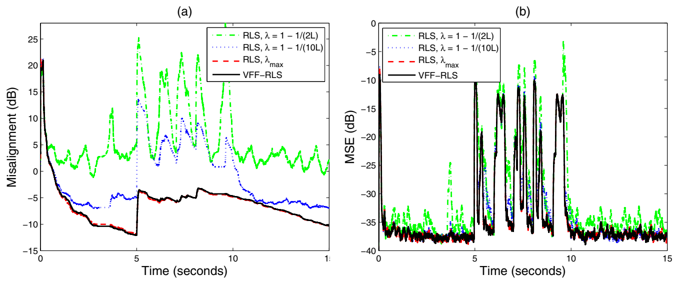

Fig. 2. Performance of the RLS algorithms using $\lambda = 1 - 1/(2L)$, $\lambda = 1 - 1/(10L)$, and $\lambda_{\text{max}} = 0.99999$, and VFF-RLS algorithm; (a) Misalignment and (b) MSE. Double-talk scenario, near-end speech appears between times 5 and 10 seconds.

near-end is a white Gaussian signal with SENR = 30 dB. The algorithms are compared in a tracking situation, i.e., the impulse responses in the near-end location are shifted to the right by 12 samples at time 7.5 seconds. The results are presented in Fig. 1. It can be noticed that the VFF-RLS algorithm tracks similarly to the RLS using the smallest forgetting factor [i.e., $\lambda = 1 - 1/(2L)$], but achieves much lower misalignment and MSE, specific to the RLS algorithm with λ_{max}. However, despite the good initial convergence rate, the tracking capabilities of the RLS algorithm using λ_{max} are reduced. Clearly, the VFF-RLS algorithm compromises much better between the previous performance criteria, by achieving a good tracking behavior but also lower misalignment/MSE.

In the second experiment, a double-talk scenario is considered, without using any double-talk detector (DTD). The near-end speech appears between times 5 and 10 seconds. The results depicted in Fig. 2 indicate that the VFF-RLS algorithm and the RLS using λ_{max} perform similarly in terms of double-talk robustness (the corresponding plots are almost overlapped), while the behavior of the RLS algorithms with the smaller forgetting factors is biased in this situation; the smaller the forgetting factor, the more biased is the algorithm during double-talk. Obviously, a DTD will improve the behavior of the algorithms in this scenario. However, it is important that the algorithm can handle a certain amount of double-talk on its own (without diverging) in order to provide good robustness features.

6. CONCLUSIONS

In this paper, we have developed a VFF-RLS algorithm in the context of SAEC with the WL model. The mechanism that controls the forgetting factor is based on the condition to recover the near-end signal from the error signal of the adaptive filter, which is always desirable in echo cancellation. The simulation results indicate that the proposed algorithm achieves good tracking capabilities but it is also robust to near-end signal variations, like double-talk. The main limitation of the VFF-RLS algorithm is its arithmetic complexity, which is similar to the RLS. However, this idea could be extended for the "fast" versions of the RLS algorithm, in order to reduce the complexity.

7. REFERENCES

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

[2] M. M. Sondhi and D. R. Morgan, "Acoustic echo cancellation for stereophonic teleconferencing," in *Proc. IEEE WASPAA*, 1991.

[3] J. Benesty, C. Paleologu, T. Gänsler, and S. Ciochină, *A Perspective on Stereophonic Acoustic Echo Cancellation*. Berlin, Germany: Springer-Verlag, 2011.

[4] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, pp. 2030–2033, Aug. 1995.

[5] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic acoustic echo cancellation–An overview of the fundamental problem," *IEEE Signal Process. Lett.*, vol. 2, pp. 148–151, Aug. 1995.

[6] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Trans. Speech, Audio Process.*, vol. 6, pp. 156–165, Mar. 1998.

[7] M. G. Bellanger, *Adaptive Filters and Signal Analysis*. NY: Dekker, 1988.

[8] P. Eneroth, S. L. Gay, T. Gänsler, and J. Benesty, "A real-time implementation of a stereophonic acoustic echo canceler," *IEEE Trans. Audio, Speech Process.*, vol. 9, pp. 513–523, July 2001.

[9] D. P. Mandic, S. Still, and S. C. Douglas, "Duality between widely linear and dual channel adaptive filtering," in *Proc. IEEE ICASSP*, 2009, pp. 1729–1732.

[10] S.-H. Leung and C. F. So, "Gradient-based variable forgetting factor RLS algorithm in time-varying environments," *IEEE Trans. Signal Process.*, vol. 53, pp. 3141–3150, Aug. 2005.

[11] C. Paleologu, J. Benesty, and S. Ciochină, "A robust variable forgetting factor recursive least-squares algorithm for system identification," *IEEE Signal Process. Lett.*, vol. 15, pp. 597–600, 2008.

[12] S. Ciochină, C. Paleologu, J. Benesty, and A. A. Enescu, "On the influence of the forgetting factor of the RLS adaptive filter in system identification," in *Proc. IEEE ISSCS*, 2009, pp. 1–4.

[13] C. Paleologu, S. Ciochină, and J. Benesty, "Double-talk robust VSS-NLMS algorithm for under-modeling acoustic echo cancellation," in *Proc. IEEE ICASSP*, 2008, pp. 245–248.