PAPER

A variable step-size adaptive cross-spectral algorithm for acoustic echo cancellation

Xiaojian LU[†a)] ***and*** **Benoit CHAMPAGNE**[†b)], ***Nonmembers***

SUMMARY The adaptive cross-spectral (ACS) technique recently introduced by Okuno *et al.* provides an attractive solution to acoustic echo cancellation (AEC) as it does not require double-talk (DT) detection. In this paper, we first introduce a generalized ACS (GACS) technique where a step-size parameter is used to control the magnitude of the incremental correction applied to the coefficient vector of the adaptive filter. Based on the study of the effects of the step-size on the GACS convergence behaviour, a new variable step-size ACS (VSS-ACS) algorithm is proposed, where the value of the step-size is commanded dynamically by a special finite state machine. Furthermore, the proposed algorithm has a new adaptation scheme to improve the initial convergence rate when the network connection is created. Experimental results show that the new VSS-ACS algorithm outperforms the original ACS in terms of a higher acoustic echo attenuation during DT periods and faster convergence rate.

key words: *acoustic echo cancellation (AEC), adaptive cross-spectral (ACS) algorithm, double-talk (DT), variable step-size, initial convergence rate*

1. Introduction

Acoustic echo cancellation (AEC) can be classified as a problem of system identification, albeit a time-varying one due to continual changes in the acoustic echo path structure. Conventional methods of adaptive system identification, e.g. least-squares algorithms [1], are affected by local disturbance signals such as the near-end speech and the background noise. When double-talk (DT) occurs, i.e. the far-end and the near-end users speak simultaneously, the adaptation of the conventional AEC device has to be stopped to avoid the possible divergence of the adaptive algorithm. Accordingly, an accurate DT detector is essential to enable a conventional AEC device to work properly. Furthermore, tracking changes in the acoustic echo path during a DT situation is particularly challenging for AEC.

During the past decade, considerable efforts have been devoted to the research of advanced AEC schemes that can properly handle DT situations. Traditionally, adaptation of the AEC device's filter coefficients is frozen by assigning a very small value (most often, 0) to the step-size of the adaptive algorithm during the DT period. In this context, many researchers have focused their efforts on developing accurate and robust DT detectors [2], [3]. Alternatively, instead of completely freezing the adaptation, various schemes with a variable step-size have been attempted to track the change in the acoustic echo path when DT occurs [4]–[7]. Unfortunately, due to the fundamental difficulties associated to distinguishing the case of DT from that of a change in acoustic echo path, most solutions are not satisfactory in terms of robustness and accuracy.

In [8], Okuno *et al.* propose an adaptive cross-spectral (ACS) algorithm that is particularly attractive for AEC applications. The ACS algorithm exploits the correlation of the far-end signal and the acoustic echo to estimate the echo path. Consequently, it has the advantage that the DT detection is not needed and that the change of the echo path can be tracked during the DT period. To achieve good levels of echo attenuation, the ACS algorithm requires the processing of many blocks of samples, so that its correlation estimate is reliable. Since the length of each block has to be larger than that of the echo path, which typically ranges from several tens up to a few hundreds of milliseconds in AEC, the ACS algorithm suffers from a relatively slow convergence rate (especially during initialization). Moreover, the non-stationary characteristics of speech signal affect the correlation estimation, leading to insufficient echo suppression during DT periods.

In this paper, a new variable step-size ACS (VSS-ACS) algorithm is proposed which can achieve faster convergence rate and a higher acoustic echo suppression in the DT situation. To this end, a generalized ACS (GACS) technique is first introduced where a step-size parameter is used to control the magnitude of the incremental correction applied to the coefficient vector of the adaptive filter. Based on the study of the effects of the step-size on the GACS convergence behaviour, a variable step-size ACS (VSS-ACS) algorithm is then proposed. To increase the convergence rate while keeping a low misadjustment, the step-size is varied dynamically by a finite state machine which monitors changes in the norm of the ACS correction applied to the adaptive filter coefficients. In addition, a new initial adaptation scheme is adopted, resulting in a significant improvement to the convergence of the algorithm at the early stage when the network connection is established. The advantages of the new algorithm are verified by computer experiments on various sets of speech files.

Manuscript received August 20, 2002.

[†]The authors are with the Department of Electrical & Computer Engineering, McGill University, 3480 University Street, Montreal, Quebec, Canada H3A 2A7

a) E-mail: xlu@tsp.ece.mcgill.ca

b) E-mail: champagne@ece.mcgill.ca

This paper is organized as follows. The generalized ACS technique is derived in Section II. The variable step-size ACS algorithm is presented in Section III along with the special scheme used to speed-up adaptation during initialization. The results and accompanying discussions of computer experiments are presented in Section IV. A brief conclusion is finally given in Section V.

2. The generalized ACS technique

The block diagram of a generic AEC system operating in the discrete-time domain is shown in Fig. 1. Let $x(n)$, $d(n)$, $\hat{y}(n)$ and $e(n)$ respectively denote the far-end speech, the microphone signal, the adaptive filter output and the residual signal at time index n.

Assume that the acoustic echo path between the loudspeaker and the microphone can be modelled as a linear, time-varying system whose impulse response at time n is represented by the N-tap vector

$$\mathbf{h}(n) = [h_0(n), h_1(n), \ldots, h_{N-1}(n)]^T \tag{1}$$

where the superscript ^{T} denotes transposition. Accordingly, the microphone signal can be expressed as

$$d(n) = \mathbf{h}(n)^T \mathbf{x}(n) + v(n) \tag{2}$$

where $\mathbf{x}(n) = [x(n), x(n-1), \ldots, x(n-N+1)]^T$ is a vector of far-end speech samples, the product $\mathbf{h}(n)^T\mathbf{x}(n)$ represents the acoustic echo, and $v(n)$ is a local disturbance signal. The latter includes both the near-end speech signal (when active) and a background noise component.

Define the coefficient vector of the adaptive filter as $\hat{\mathbf{h}}(n)$, whose length is also assumed to be N taps, so that the adaptive filter output can be expressed as

$$\hat{y}(n) = \hat{\mathbf{h}}(n)^T \mathbf{x}(n) \tag{3}$$

The residual signal sent to the far-end user after echo cancellation is thus given by

$$\begin{aligned} e(n) &= d(n) - \hat{y}(n) \\ &= d(n) - \hat{\mathbf{h}}(n)^T \mathbf{x}(n) \end{aligned} \tag{4}$$

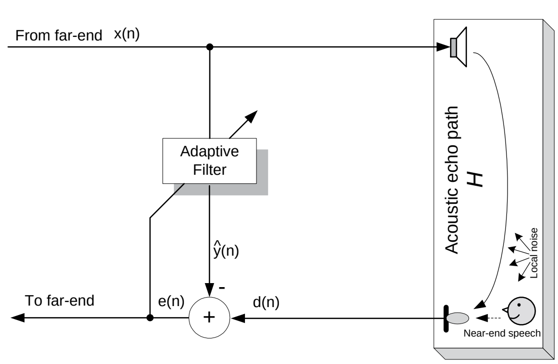

Fig. 1 Block diagram of an AEC system.

Partitioning the data into consecutive blocks of length M samples and taking the short-term Fourier transform (stDFT) [9] of length $K \geq M$ (zero-padding assumed) on both sides of Eq. (4), the residual signal is expressed in the frequency domain as

$$E(k;m) = D(k;m) - \hat{H}(k;m)X(k;m) \tag{5}$$

where $E(k;m)$, $D(k;m)$ and $X(k;m)$ respectively denote the stDFTs of the signals $e(n)$, $d(n)$ and $x(n)$ over one block, while $\hat{H}(k;m)$ is the DFT of the vector $\hat{\mathbf{h}}(n)$, assumed to remain constant within the duration of a block. Note that $\hat{H}(k;m)$ represents the estimated frequency response of the echo path. In Eq. (5), the parameter $k \in \{1, 2, \ldots, K\}$ is the index of the frequency bin and $m \in \{1, 2, \ldots\}$ is the block index in the time domain. For the linear convolution in Eq. (4) to be equivalent to the circular convolution in Eq. (5), the stDFT size K should be such that $K \geq M + N - 1$.

Define the cost function

$$J_k = \mathrm{E}[|E(k;m)|^2] = \mathrm{E}[E(k;m)E^*(k;m)] \tag{6}$$

where $\mathrm{E}[\cdot]$ denotes statistical expectation. The update equation of the generalized ACS technique is obtained via the steepest descent iterative philosophy [10] applied to the cost function J_k, namely:

$$\hat{H}(k;m+1) = \hat{H}(k;m) - \frac{\mu'}{2}\frac{\partial \mathrm{E}[|E(k;m)|^2]}{\partial \hat{H}(k;m)} \tag{7}$$

In the frequency domain, the signals and the adaptive filter coefficients are generally complex valued. Splitting the various quantities in Eq. (5) into real and imaginary components one easily obtains

$$\begin{aligned} E(k;m) = \ & D_R(k;m) - \hat{H}_R(k;m)X_R(k;m) \\ & +\hat{H}_I(k;m)X_I(k;m) \\ & +j[D_I(k;m) - \hat{H}_R(k;m)D_I(k;m) \\ & -\hat{H}_I(k;m)X_R(k;m)] \end{aligned} \tag{8}$$

where the subscripts R and I are used to denote the real and imaginary parts of the corresponding quantity, respectively. The partial derivative in Eq. (7) can then be computed as [10]

$$\begin{aligned} \frac{\partial \mathrm{E}[|E(k;m)|^2]}{\partial \hat{H}(k;m)} &= \mathrm{E}\left[\frac{\partial |E(k;m)|^2}{\partial \hat{H}_R(k;m)} + j\frac{\partial |E(k;m)|^2}{\partial \hat{H}_I(k;m)}\right] \\ &= -2\mathrm{E}[X^*(k;m)E(k;m)] \end{aligned} \tag{9}$$

Applying an NLMS-like procedure of normalization, i.e.

$$\mu' = \frac{\mu}{\mathrm{E}[|X(k;m)|^2]} \tag{10}$$

the adaptive filter update shown in Eq. (7) can be written as

$$\hat{H}(k;m+1) = \hat{H}(k;m) + \mu\Delta H(k;m) \tag{11}$$

$$\Delta H(k;m) = \frac{\mathrm{E}[X^*(k;m)E(k;m)]}{\mathrm{E}[|X(k;m)|^2]} \tag{12}$$

We note that Eq. (12) can be expanded as

$$\Delta H(k;m) = H(k;m) - \hat{H}(k;m) + \frac{\mathrm{E}[X^*(k;m)V(k;m)]}{\mathrm{E}[|X(k;m)|^2]} \tag{13}$$

where $H(k;m)$ is the frequency response of the true (unknown) acoustic echo path and $V(k;m)$ is the stDFT of the local disturbance signal $v(n)$. Clearly, during a DT situation, under the assumption that the far-end signal $x(n)$ is uncorrelated with $v(n)$, or at least the correlation between these signals is very weak, the second term in Eq. (13) can be neglected. Accordingly, the local DT signal will have no effect on the algorithm.

The main difficulty associated to the implementation of Eq. (12) is that it requires prior knowledge of the signals' second order statistics. While such knowledge is hardly available in most cases, the desired expectations can be estimated in practice by means of time averaging over the block index m. Among the various types of sliding windows which can be employed for online time averaging, the most commonly used are the exponential and the rectangular ones, with varying degrees of temporal overlap. However, as our experience indicates, a very small step-size has to be used with either types of windows when there is significant overlap between successive windows, which in turns results in a slower convergence rate. For example, in the case of a rectangular sliding window with length L ($L \gg 1$) blocks and a minimum shift of one block between each update of the time averages, the step-size in Eq. (11) must be significantly reduced to avoid divergence of the algorithm due to the strong correlation between successive gradient update directions.

Here, we assume a rectangular sliding window of length L, with a window shift of Q ($1 \le Q \le L$) blocks for each update of the algorithm. Accordingly, the estimated echo path is updated once every Q blocks, with the incremental correction computed as follows:

$$\Delta H(k;p) = \frac{\displaystyle\sum_{i=(p-1)Q+1}^{(p-1)Q+L} X^*(k;i)E(k;i)}{\displaystyle\sum_{i=(p-1)Q+1}^{(p-1)Q+L} |X(k;i)|^2} \tag{14}$$

where, $k \in \{1,2,\ldots,K\}$, and $p \in \{1,2,\ldots\}$.

To avoid any processing delay introduced to the algorithm, the operation of the adaptive filter is preferably carried out in the time domain, except for the computation of $\Delta H(k;p)$ in Eq. (14). Thus, the adaptive filter coefficient vector is updated every MQ samples in the time domain, through:

$$\hat{\mathbf{h}}_{p+1} = \hat{\mathbf{h}}_p + \mu\Delta\mathbf{h}_p. \tag{15}$$

where $\Delta\mathbf{h}_p$ is the inverse DFT of $\Delta H(k;p)$. The relationship between the iteration index p and the sample index n is

$$p = \left\lceil \frac{n}{MQ} \right\rceil, \quad n \in \{1,2,\ldots\} \tag{16}$$

where, $\lceil\delta\rceil$ represents the smallest integer $\ge \delta$. Therefore, the coefficient vector of the adaptive filter $\hat{\mathbf{h}}(n)$ remains constant for MQ consecutive samples.

We shall refer to Eqs. (4), (14) and (15), together with the associated stDFT computations, as the generalized ACS (GACS) algorithm. The original ACS [8] is a special case of GACS when μ is set to 1 in Eq. (15). The introduction of the step-size in GACS will allow additional flexibility in implementation for improved performance (see next Section). It is noted that the step-size μ must be carefully chosen to ensure the convergence of the algorithm, i.e. $0 \le \mu \le \mu_{\max}$. An exact theoretical expression for the upper bound $\mu_{\max}$ cannot be derived easily under practical conditions of operation with non-stationary, non-white inputs (e.g. speech signals). Experimentally, we have observed that $\mu_{\max}$ depends on the type of input signal, the rectangular window length L and the data reuse factor (i.e. window overlap ratio) $(L-Q)/L$. In all of the practical cases that we have tested, the GACS algorithm worked properly (i.e. no observed instability) when μ was selected within the range $0 \le \mu \le 1.5$.

We note that there is a certain degree of similarity between GACS and the standard frequency domain adaptive filter (FDAF) [10]. However, although both GACS and FDAF are operated in the frequency domain, the former uses block averaging to estimate the required expected values, while the later uses instantaneous estimation (i.e. a single block). Accordingly, GACS is much more robust than FDAF to disturbance signals, as long as a proper number of blocks is employed in the time averaging operation in Eq. (14).

Finally, the behaviour of GACS as a function of its step-size μ is consistent with that of other steepest descent adaptive filters (including FDAF). That is, increasing the step-size results in a faster convergence rate at the expense of higher coefficient misadjustment in the steady-state. (See Section 4.2 for results).

3. Variable step-size ACS

3.1 Finite state machine

The observed performance of GACS suggests that the use of a variable step-size could improve the convergence rate and reduce the misadjustment. Specifically, a larger value of μ should be used during acoustic echo path changes and a smaller value should be used when the algorithm has converged. Although many adaptive filtering algorithms with variable step-sizes have been studied, there is still a lack of a robust algorithm which can accurately distinguish an echo path change from a DT situation and as well, track the echo path change during a DT period.

Based on the characteristics of the GACS algorithm, a variable step-size ACS (VSS-ACS) mechanism is now proposed. The latter consists of the GACS algorithm, accompanied by a step-size adjustment mechanism that is regulated by a finite state machine as shown in Fig. 2. Three regions of operation are identified in the figure that correspond to different states, as explained below:

Region I: This region corresponds to a fast tracking mode where the algorithm needs to estimate or track large changes in the acoustic echo path. Initial operation of the algorithm upon network connection also falls into this category. In order to quickly track the changes in the echo path, a larger step-size, denoted as μ_I, is necessary.

Region II: This region represents a transient mode of the algorithm. After rapid adaptation in Region I, misadjustment (due to gradient noise) gradually becomes an issue while it is still important to keep the adaptive filter updating its weight vector at a reasonable rate. The introduction of two states with smaller step-sizes μ_{IIa} and μ_{IIb} in this region ensures the smooth and robust transition from Region I to Region III (see below).

Region III: Corresponding to the steady-state of the algorithm, a small step-size μ_{III} is proper in this region. Two factors actually limit the choice of the step-size. On the one hand, the step-size should not be too small so as to prevent the VSS-ACS to track the small, ever present fluctuations in the acoustic echo path. On the other hand, it should not be too large so as to minimize misadjustment and to allow the estimated echo path error to increase during DT periods. Indeed, while in theory the GACS is not affected by DT, it will be in a practical implementation of Eq. (12) based on temporal averaging with non-stationary signals.

The adaptive filter coefficient error could be used in theory to determine the proper state of the algorithm. Unfortunately, it is almost impossible to obtain this information in practice because the acoustic echo path is unknown and time-varying. As an alternative, we propose to use the energy ratio of two successive increments of the estimated system impulse response as the basis of a series of tests to decide upon state transition. Denoted as $\delta_h(p)$, this energy ratio is formally defined here as

$$\delta_h(p) = \left| 20 \log_{10} \frac{\|\Delta \hat{\mathbf{h}}_p\|}{\|\Delta \hat{\mathbf{h}}_{p-1}\| + c} \right| \tag{17}$$

where, $\|\mathbf{u}\|$ is the norm of the vector $\mathbf{u}$ and c is a small constant preventing the division from overflow. As our experiments indicate, this energy ratio is strongly indicative of the true (but practically unknown) error in the adaptive filter coefficients.

Referring to Fig. 2, four states have been identified for the VSS-ACS algorithm. Transitions between these states are determined on the basis of the comparison between $\delta_h(p)$ and a set of thresholds, labelled $\lambda_1, \lambda_2, \ldots, \lambda_5$. For example, when in the state μ_{IIa}: a transition to state μ_{IIb} will occur if $\delta_h(p) < \lambda_2$; a transition to state μ_I will occur if $\delta_h(p) > \lambda_5$; else the algorithm will remain in state μ_{IIb}, as indicated by the self-loop. Additional guidelines and considerations in the selection of the thresholds are discussed below.

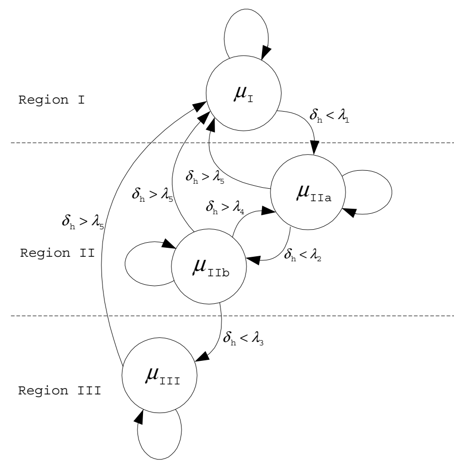

Fig. 2 State machine diagram of the VSS-ACS algorithm.

First consider the case where the acoustic environment does not change significantly over time. During the adaptation of VSS-ACS, the state of the algorithm is then expected to transit from Region I to Region II, or from Region II to Region III. Here, it is quite straightforward to set the values of λ_1, λ_2 and λ_3, which are in descending order, representing different stages on the way of convergence of the algorithm.

Next, consider a situation where the acoustic echo path is notably changed (including initialization). In this case, the algorithm is expected to track the variation as fast as possible. Hence, the state of the algorithm should jump to Region I, no matter which region it previously stayed. One major advantage of the GACS technique over more traditional adaptive filtering approaches is that $\delta_h(p)$ will not increase significantly in the DT situation, while a sudden significant change in the echo path will manifest itself by a corresponding large change in $\delta_h(p)$. Thus, the VSS-ACS has the intrinsical ability to distinguish between echo path change and DT situation. Accordingly, the threshold for the change of the acoustic echo path, λ_5, is significantly higher than others.

Finally, in Region II, two states are introduced for the improved robustness of the algorithm. This is because the adaptive filter coefficient update may be unfavourably affected by a specific segment of speech input. In this case, more time is needed to decide whether or not the algorithm has reached the steady-state where the smallest step-size μ_{III} can be used. Based on this consideration, it is reasonable to set $\lambda_4 \geq \lambda_2$.

Specific numerical values for the various parameter of the VSS-ACS algorithm, i.e. step-sizes and thresholds, will be given in Section 4.1.

3.2 Improvement to initial convergence

Among the total LM samples used to compute the incremental correction in Eq. (14), only MQ samples are new data. Thus, similar to the original ACS, VSS-ACS updates the estimated echo path every MQ samples. At the beginning of adaptation, when the network connection has just been created, the initial values of the adaptive filter coefficients are set to 0, and the acoustic echo is not attenuated until ACS has collected at least QM samples (assuming $(L-Q)M$ padding zeros) to estimate the echo path. In AEC, the acoustic echo path may be as long as several hundred or even thousand taps at 8kHz sampling rate. Consequently, with Q sufficiently large, say 40 or more [8], and $M \geq N$, where N is the adaptive filter length, QM samples represent a relatively long time interval. It is obviously inappropriate that the acoustic echo is not suppressed during such a long initial period. In order to overcome this drawback, a modification of the VSS-ACS algorithm is described below where different approximation and adaptation schemes are used to computation the filter weight vector during the initial period. The proposed approach is an extension of the technique originally reported in [11] for the ACS algorithm.

Referring to Eq. (2), the microphone signal in the frequency domain may be expressed as

$$D(k;m) = H(k;m)X(k;m) + V(k;m) \tag{18}$$

where $H(k;m)$ and $V(k;m)$ have been previously defined in connection with Eq. (13). Multiplying both sides of Eq. (18) by $X^*(k;m)$, taking the expectation and assuming that $x(n)$ and $v(n)$ are uncorrelated, the frequency response of the acoustic echo path is easily obtained as

$$H(k;m) = \frac{\mathrm{E}[X^*(k;m)D(k;m)]}{\mathrm{E}[|X(k;m)|^2]} \tag{19}$$

corresponding to the so-called cross-spectral technique [12].

Similarly to the approximation process leading from Eq. (12) to Eq. (14), the expected values in Eq. (19) can be estimated by time averaging. However, to make full use of the available data so as to improve the initial convergence rate, the desired expected values are estimated by accumulators running from block 1 up to current block m (instead of the summations over L blocks as in Eq. (14)). Therefore, the initial adaptation is expressed as

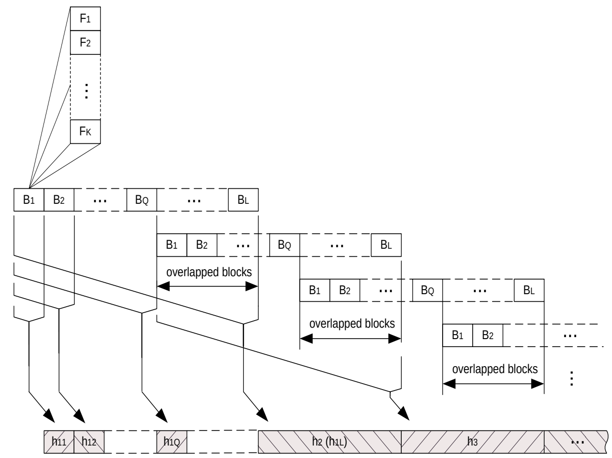

Fig. 3 Adaptation of the VSS-ACS algorithm.

$$\hat{H}(k;m) = \frac{\sum_{i=1}^{m} X^*(k;i)D(k;i)}{\sum_{i=1}^{m} |X(k;i)|^2} \tag{20}$$

where the block index $m \in \{1, 2, \ldots, L\}$. The coefficients of the estimated acoustic echo path are then obtained by taking inverse DFT on Eq. (20). To reduce the systematic error introduced by the cross-spectral technique, i.e. the magnitude of the estimated echo path impulse response being smaller than that of the real one due to the so-called deformation phenomenon [12], the estimated value needs to be enlarged by a multiplying factor. For consistency, this factor is also denoted by μ except that here, $\mu \geq 1$. Hence, for initialization, the adaptive filter coefficient vector is updated every M samples via

$$\hat{\mathbf{h}}_{1m} = \text{inverse DFT}\{\mu \hat{H}(k;m)\} \tag{21}$$

It is noted that the coefficients of the estimated echo path are computed directly in the initial period, while they are computed recursively in the subsequent periods.

The blocking procedure of the VSS-ACS algorithm with modified initialization is illustrated in Fig. 3, where B_i refers to a particular block of m samples and F_k to a particular frequency bin within a block.

3.3 Computational complexity

To efficiently implement the VSS-ACS algorithm, an FFT algorithm can be employed for the K-point stDFT computations. Accordingly, only $K \log_2 K$ operations [13] are needed to map each data block from the time domain to the frequency domain, and vice versa, where the term *operation* refers to one multiplication and one addition. The computational complexities for Eqs. (4), (14) and (15) are about LMN, $(L+Q)K$

and N operations per adaptive filter coefficient update, respectively.

Referring to the state machine, the logarithmic scale is used in the definition of $\delta_h(p)$ in Eq. (17) only to simplify understanding. In practice, both the values of the thresholds and the energy ratio $\delta_h(p)$ can be calculated in the linear scale, so that the computational complexity is reduced. Hence, only the norm of the current coefficient increment needs to be computed, with a computational load N operations per filter coefficient update, approximately.

During the initial period, the practical implementation can exploit recursive relations to reduce the computational burden. Specifically, $\hat{H}(k;m)$ in Eq. (20) can be computed as

$$\begin{aligned} \hat{H}(k;m) &= \frac{P(k;m)}{Q(k;m)} \\ P(k;m) &= P(k;m-1) + X^*(k;m)D(k;m) \\ Q(k;m) &= Q(k;m-1) + |X(k;m)|^2 \end{aligned} \quad (22)$$

where $P(k;m)$ and $Q(k;m)$ respectively stand for $\sum_{i=1}^{m} X^*(k;i)D(k;i)$ and $\sum_{i=1}^{m} |X(k;i)|^2$. Accordingly, the computational complexity of the modified initialization scheme is obtained as $MN + N + 3K\log_2 K$ operations per coefficient update.

The computational complexity of the VSS-ACS algorithm, in common units of operations per sample (OPS), is therefore obtained as follows:

a) During the initial period

During this period, the algorithm updates the adaptive filter coefficient vector every M samples. Consequently, the computational complexity is

$$\frac{MN + N + 3K\log_2 K}{M} \text{ OPS} \quad (23)$$

For the case that the block length is the same as that of the adaptive filter, i.e. $M = N$, and $K = M + N$ for linear convolution, the computational complexity in the initial period is about $N + 6\log_2 N$ OPS.

b) In the subsequent period

Because the filter coefficients are updated every QM samples, the total computational complexity of the VSS-ACS algorithm is

$$\frac{LMN + (L+Q)K + 2N + (L+Q)K\log_2 K}{MQ} \text{ OPS} \quad (24)$$

For the special case $M = N$ and $K = 2N$, and assuming one half window overlap, i.e. $Q = L/2$ blocks in Eq. (14), the computational complexity is about $2N + 6(\log_2 N)/N$ OPS.

4. Computer experiments

4.1 Methodology

In the computer experiments, various segments of speech signals including males', females' and children's speech were used as the excitation signals. A coloured noise, obtained by passing a white noise signal through a first-order IIR filter $H(z) = 1/(1-0.9z^{-1})$, was added into the microphone signal to simulate the local noisy environment, so that the echo-to-noise ratio (ENR) is 30dB. The impulse response of the acoustic echo path, whose length is $N = 300$ taps, corresponding to 37.5ms at 8kHz sampling rate, was synthesized to mimic the driver's compartment of a motor vehicle.

To evaluate the performance of the algorithms, the following normalized measure of coefficient error of the estimated impulse response at time n was introduced (dB unit):

$$\text{Coef\_err}(n) = 20\log_{10}\left(\frac{\|\mathbf{h}(n) - \hat{\mathbf{h}}(n)\|}{\|\mathbf{h}(n)\|}\right) \quad (25)$$

As mentioned earlier, $\mathbf{h}(n)$ and $\hat{\mathbf{h}}(n)$ respectively represent the coefficient vector of the true acoustic echo path and that of the estimated one.

In the implementation of the GACS and VSS-ACS algorithms, the block size was $M = 300$ samples, $L =$ 80 blocks were used for the average in Eq. (14), and the window shift between updates was set to $Q = 40$ blocks. For the VSS-ACS algorithms, the set of step-sizes in the three regions were: $\mu_I = 1.2$, $\mu_{IIa} = 1.0$, $\mu_{IIb} = 0.8$ and $\mu_{III} = 0.1$. Correspondingly, the thresholds of the state machine were 5.5, 3.0, 1.5, 4.0, 20.0 for $\lambda_1, \ldots, \lambda_5$, respectively.

The properties of the VSS-ACS algorithm have been tested for three different aspects, namely: behaviour in the DT situation, initial convergence rate, and tracking capability in the presence of the near-end speech. For comparison, these tests are also applied to the original ACS algorithm, with same values of M, L and Q as above.

4.2 Effect of the step-size μ on GACS

To support our previous statement regarding the effect of the step-size μ on the GACS convergence rate, we present the time evolution of the coefficient error with various step-sizes in Fig. 4. The adaptive filter weight vector was initially set to zero; the excitation signal was speech. It can be seen that increasing the step-size μ in the GACS algorithm leads to faster initial convergence, but higher misadjustment in the steady-state, as pointed out earlier.

We note that GACS inherits the robustness properties of the original ACS algorithm to local disturbance

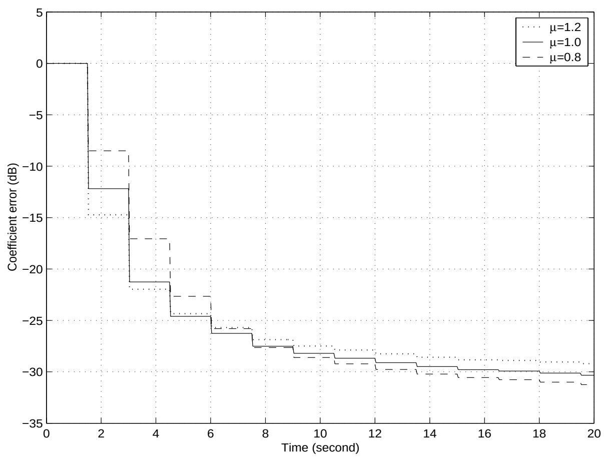

Fig. 4 Effects of the step-size on the GACS algorithm.

signals. Indeed, while ENR was set to 30dB in Fig. 4, results of other experiments show that the algorithm convergence behaviour is essentially unaffected when the power of the disturbance signal, e.g. additive noise, varies in a large range, say for ENR as low as 5 to 10dB.

4.3 Behaviour of VSS-ACS during DT

In this experiment, a segment of speech signal which had a comparable power with the acoustic echo was added into the microphone signal to simulate the DT situation. Moreover, in order to clearly demonstrate the effect of DT on the algorithms, DT occurred after initial convergence of the algorithms.

Figure 5 shows the performance of the original ACS algorithm (i.e. GACS with $\mu = 1$) and the VSS-ACS algorithm during the DT period. The microphone signal, whose waveform is displayed in Fig. 5(b), comprises the acoustic echo and the near-end signal. The latter consists of a near-end speech superimposed on a background noise, as plotted in Fig. 5(a). The signal waveforms of the residual echo produced by ACS and VSS-ACS are shown in Fig. 5(c) and 5(d), respectively. Note that the near-end signal has been subtracted from the residual signal for clarity. These waveforms demonstrate that VSS-ACS provides much more attenuation to the acoustic echo than ACS during a DT situation.

The corresponding errors of the estimated coefficients of the acoustic echo path have also been compared in Fig. 6 for the proposed VSS-ACS algorithm and the original ACS algorithm. It reveals that the new VSS-ACS algorithm has notably smaller coefficient error than ACS in the DT situation, which agrees with the results presented in Fig. 5.

4.4 Improvement to initial convergence

Here, the initial convergence properties of different schemes are examined. The signal waveforms of the

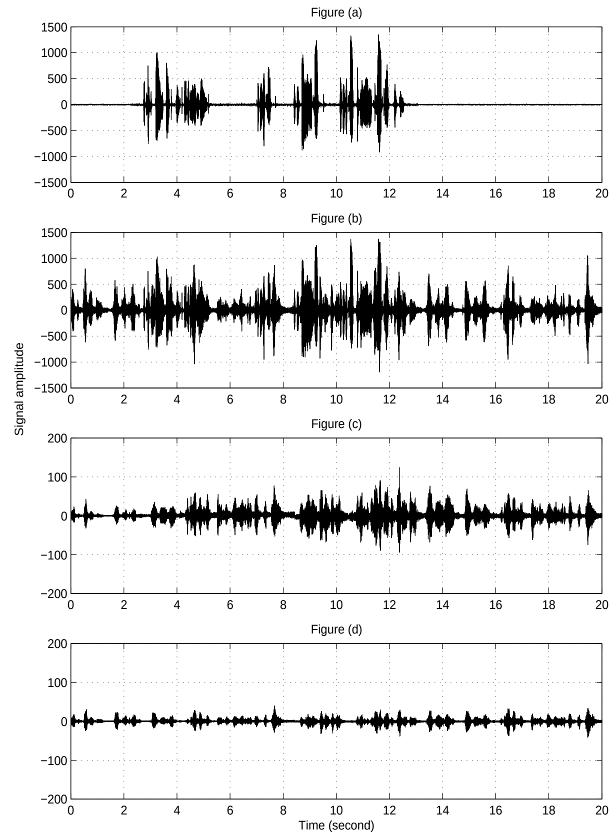

Fig. 5 Waveforms in the DT situation: (a) original near-end signal; (b) microphone signal (acoustic echo plus near-end signal); (c) residual echo of ACS; (d) residual echo of VSS-ACS.

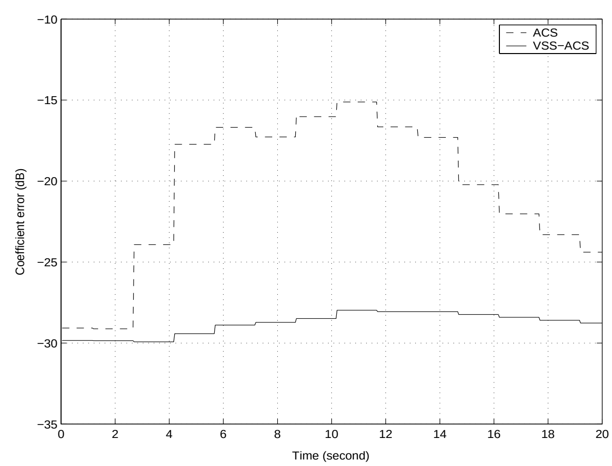

Fig. 6 Coefficient errors versus time for the original ACS algorithm (dash line) and VSS-ACS algorithm (solid line).

acoustic echo, residual echoes of ACS and VSS-ACS are shown in Fig. 7(a), 7(b) and 7(c), respectively. It is obvious that the VSS-ACS algorithm achieves a much faster convergence rate when the network connection is created. The time evolution of coefficient errors of the echo path for both ACS and VSS-ACS are plotted in Fig. 8. Clearly, VSS-ACS shows a significant improve-

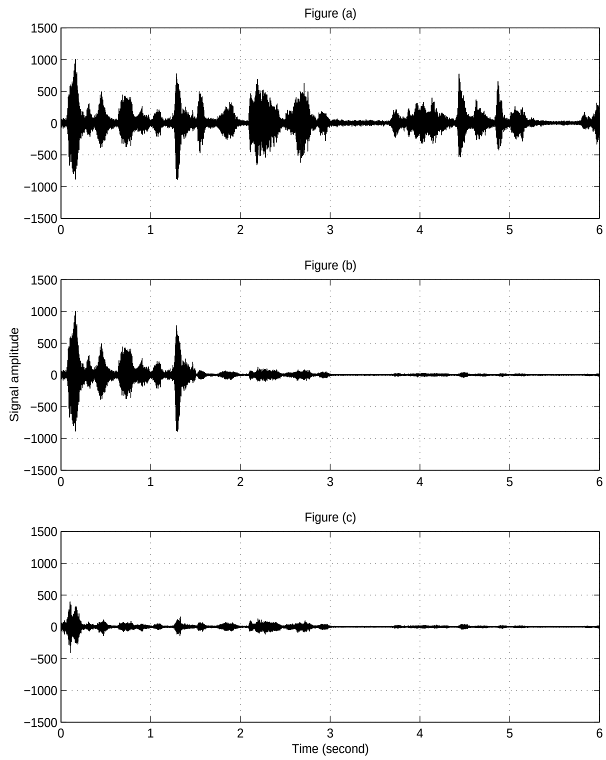

Fig. 7 Waveforms in the initial period: (a) acoustic echo signal; (b) residual echo of ACS; (c) residual echo of VSS-ACS.

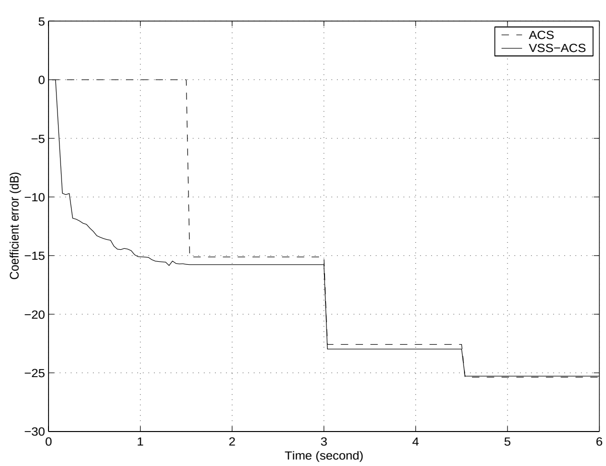

Fig. 8 Coefficient errors versus time in the initial period: dash line – ACS; solid line – VSS-ACS.

ment in terms of acoustic echo suppression during the initial period as a result of the use of the modified initialization scheme put forward in Section 3.2.

4.5 Tracking in the presence of DT

In order to test the tracking capability of the new algorithm, the acoustic echo path is changed from $\mathbf{h}(n)$ to $-\mathbf{h}(n)$ during DT after the algorithm has converged. The coefficient errors for the original ACS and VSS-ACS are displayed in Fig. 9. It is observed that the proposed algorithm not only inherits the merit of ACS which can track the change of the acoustic echo path in the presence of strong disturbance signal, e.g. the DT situation, but also presents a better performance.

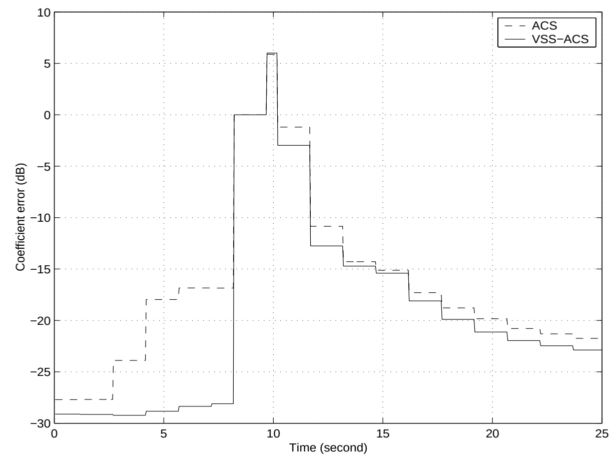

Fig. 9 Coefficient errors versus time (acoustic echo path changed from $\mathbf{h}(n)$ to -$\mathbf{h}(n)$) for the original ACS (dash line) and VSS-ACS (solid line).

4.6 Subjective experiments

The results of informal listening tests suggest that, compared to the original ACS algorithm, the proposed VSS-ACS can suppress the acoustic echo to a satisfactory level even during initialization. In DT situations, the near-end speech contained in the residual signal, which is sent to the far-end user, is more clearly audible with the VSS-ACS than with ACS because of the lower level of interference signal, i.e. residual acoustic echo. Furthermore, no perceptual distortion of the near-end speech signal is observed. In the case when the acoustic echo path changes, the acoustic echo is suppressed more rapidly by VSS-ACS.

5. Summary and conclusion

A generalized ACS technique was proposed where a step-size parameter is used to control the magnitude of the incremental correction applied to the coefficient vector of the adaptive filter. Based on the study of the effects of the step-size on the ACS convergence behaviour, a new variable step-size ACS (VSS-ACS) algorithm is proposed, where the value of the step-size is commanded dynamically by a special finite state machine, so as to optimize algorithm performance in terms of convergence rate and misadjustment in various situations. Furthermore, the proposed algorithm has a new adaptation scheme to improve the initial convergence rate when the network connection is created.

The proposed VSS-ACS algorithm is attractive in the application of AEC because it can remarkably attenuate acoustic echo even in the presence of near-end speech and high level background noise, without the requirement of DT detection, which is still a difficult issue. The computational complexity of the VSS-ACS algorithm is comparable to that of the standard LMS algorithm, allowing for low-cost real-time implementation with existing DSP technology. The results of computer experiments show that the new VSS-ACS algorithm outperforms the original ACS, both in terms of superior acoustic echo suppression during DT periods and faster initial convergence.

Acknowledgement

This work has been supported in part by the Natural Sciences and Engineering Research Council of Canada.

References

[1] G. Glentis, K. Berberidis and S. Theodoridis, "A unified view: efficient least squares adaptive algorithms for FIR transversal filtering," IEEE Signal Processing Magazine, vol. 16, no. 4, pp. 13–41, Jul., 1999.

[2] S.H. Kim, H.S. Kwon, K.S. Bae, K.J. Byun and K.S. Kim, "Performance improvement of double-talk detection algorithm in the acoustic echo canceller," Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, vol. 5, pp. 3249–3252, 2001.

[3] J.H. Cho, D.R. Morgan and J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," IEEE Trans. Speech and Audio Processing, vol. 7, no. 6, pp. 718–724, Nov. 1999.

[4] C. Breining, "A robust fuzzy logic-based step-gain control for adaptive filters in acoustic echo cancellation," IEEE Trans. Speech and Audio Processing, vol. 9, no. 2, pp. 162–167, Feb. 2001.

[5] S.J. Pushparajah and J.A. Chambers, "A smarter method for acoustic echo cancellation in the presence of double talk," Proc. IEEE Workshop on Acoustic Echo and Noise Control, pp. 184–186, 1999.

[6] P. Heitkämper, "An adaptation control for acoustic echo cancellers," IEEE Signal Processing Letters, vol. 4, no. 6, pp. 170–172, Jun. 1997.

[7] Q.-G. Liu and B. Champagne, "Adaptive echo cancelling system for telephony applications," United States Patent 6108412, Aug. 2000.

[8] T. Okuno, M. Fukushima and M. Tohyama, "Adaptive cross-spectral technique for acoustic echo cancellation," IEICE Trans. Fundamentals, vol. E82-A, no. 4, pp. 634–639, Apr., 1999.

[9] J. R. Deller, J. G. Proakis and J. H. L. Hansen, Discrete-time processing of speech signals, IEEE Press, New York, 2000.

[10] P. M. Clarkson, Optimal and Adaptive Signal Processing, CRC Press, 1993.

[11] X. Lu and B. Champagne, "Acoustic echo cancellation over a non-linear channel," Proc. IEEE Workshop on Acoustic Echo and Noise Control, pp. 139–142, 2001.

[12] M. Fukushima, T. Okuno, H. Yanagawa and K. Kido, "Improvement of the accuracy in attenuation constant estimation using the cross-spectral technique," IEICE Trans. Fundamentals, vol. E82-A, no. 4, pp. 626–633, Apr., 1999.

[13] A. V. Oppenheim, R. W. Schafer and J. R. Buck, Discrete-time signal processing, Prentice Hall, NJ, 1998.