A VARIABLE STEP-SIZE CONTROL OF ADAPTATION FOR ACOUSTIC ECHO CANCELLATION FILTERS

Vojko Pahor *Alberto Carini*

T́elit Mobile Terminals S.p.A.
Viale Stazione di Prosecco 5/B
I-34010 Sgonico (Trieste) Italy
E-Mails: `vojko.pahor, alberto.carini@telital.com`

ABSTRACT

This paper presents a low-complexity variable step-size control for acoustic echo cancellation filters. The proposed control method is based on two supplementary adaptive filters with low memory length. The first supplementary filter is used to control the double-talk situations while the second supplementary filter is employed to detect variations of the echo-path impulse response. The variable step-size here derived is independent from the echo coupling loss and from the reverberation time and thus it is suitable for both hands-free and handset terminals.

1. INTRODUCTION

The acoustic echo is an annoying disturbance which affects both hands-free and handset telephone terminals. The disturbance is originated by the sound propagation from the loudspeaker to the microphone of the receiver. The problem is particularly relevant in digital cellular telephony because of the long propagation delay introduced by the telecommunication system. In this area, acoustic echo cancellers are typically employed in order to cope with the disturbance. A commercial acoustic echo canceller is constituted by several components that contributes all to the success or to the failure of the device. Indeed, the high correlation of the input signal (*i.e.* the far-end speech), the long reverberation time of the enclosure, the highly time-varying characteristics of the echo path, the problem of the local disturbance (*i.e.* of the local noise and, more important, of the local speech), make really challenging the development of an effective acoustic echo cancellation system. The main components of a commercial acoustic echo canceller are: 1) the digital adaptive filter that estimates the acoustic echo in order to cancel it by subtraction; 2) the adaptation control that copes with the local disturbance and with the echo impulse response variations; 3) the residual echo suppresser that eventually provides the desired target echo attenuation. Despite several contributions can be found in literature on different filter adaptation techniques suitable for acoustic echo cancellation (see references [1] and [2]), the problems of the canceller adaptation control and of the residual echo control have been for long time underestimated or ignored and only in recent years they have raised a true interest of researchers. In this paper we deal with the problem of the echo canceller adaptation control. Different control algorithms have been proposed in literature (see references [3], [4] and the references there defined). Some adaptation control methods are implemented by switching the step-size of the adaptive filter between some fixed values. These adaptation controls usually employ detectors based on measurements of signal powers [5] or of signal autocorrelations [6], [7] in order to choose one of the possible values of the step-size. Other adaptation controls directly compute at each filter update a suitable value of the step-size. These methods are called variable step-size methods and they typically try to estimate the optimal step-size [4], [8], which depends on the power of a non-measurable signal. Therefore, any variable step-size method can only approximately estimate the optimal step-size from some measurements of the far-end speech, of the near-end speech, of the local noise and of the system error. Most variable step-size methods proposed in literature present some limitations [4]. For instance, some of these methods require an accurate detector of the local speech activity, some others are not able to discriminate the local speech form an echo-path variation and they require a rescue detector.

In this paper we present a low complexity control for acoustic echo cancellers which efficiently copes with both double-talk and echo variations. The adaptation control is based on a novel variable step-size formula which employs two short supplementary adaptive filters to monitor the near-end activity and the echo variations. Unlike the methods based on power or correlation measures comparisons, the proposed method does not need any estimate of the echo coupling loss and thus it can operate efficiently in any environment (*e.g.* car, cellular phone, office).

The rest of this paper is organized as follows. Section 2 presents and discusses the novel adaptation control. Section 3 discusses an experimental result obtained in a car hands-free system. Conclusions follow in Section 4.

2. THE ADAPTATION CONTROL

In this section we present the adaptation control for the NLMS adaptive algorithm. Our derivation can be repeated for any gradient descendent adaptive filter.

Throughout the paper we refer to the system of Figure 1 where: $x(n)$ indicates the far-end speech signal, $c(n)$ the acoustic echo, $s(n)$ the local speech, $v(n)$ the local noise, $y(n)$ the output of the acoustic echo canceller, $d(n) = c(n) + s(n) + v(n)$ the microphone signal and $e(n) = d(n) - y(n)$ the echo cancelled signal.

The acoustic echo canceller estimates the impulse response of the loudspeaker-enclosure-microphone system. Due to the analog-to-digital and the digital-to-analog conversions and to the fly-time between the loudspeaker and the microphone, the acoustic echo $c(n)$ is a delayed version of the far-end signal $x(n)$. Therefore, the acoustic echo can be estimated with a filter $\mathbf{w}_B(n)$ from the delayed signal $x(n - N_T)$ where N_T is less or equal to the delay introduced by the system (if necessary, the delay of N_T taps can be artificially introduced). Our adaptation control is based on two supplementary filters, $\mathbf{w}_A(n)$ and $\mathbf{w}_C(n)$, that process the most N_T recent samples of $x(n)$. These two filters estimate the local speech activity and detect any variation of the echo impulse. Equations (1), (2) and (3) give the input-output relationships of the acoustic echo cancellation filter $\mathbf{w}_B(n)$, and of the two supplementary filters $\mathbf{w}_A(n)$ and $\mathbf{w}_C(n)$, respectively.

$$y(n) = y_B(n) = \sum_{k=0}^{N-N_T-1} w_B(n)[k] \cdot x(n-N_T-k), \quad (1)$$

$$y_A(n) = \sum_{k=0}^{N_T-1} w_A(n)[k] \cdot x(n-k), \quad (2)$$

$$y_C(n) = \sum_{k=0}^{N_T-1} w_C(n)[k] \cdot x(n-k). \quad (3)$$

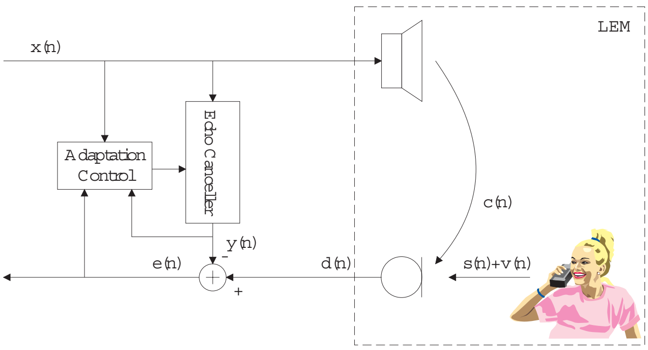

Figure 1: The acoustic echo canceller system.

In equations (1), (2) and (3) the notation $w_H(n)[k]$ indicates the k-th coefficient of the generic filter $\mathbf{w}_H$ at time n.

The NLMS update equations for the coefficients of the three filters are given in equations (4), (5) and (6),

$$\mathbf{w}_B(n+1) = \mathbf{w}_B(n) + \frac{\mu_B(n)}{\mathbf{x}^T(n)\mathbf{x}(n)} e_B(n)\mathbf{x}_B(n), \quad (4)$$

$$\mathbf{w}_A(n+1) = \mathbf{w}_A(n) + \frac{\mu_A(n)}{\mathbf{x}^T(n)\mathbf{x}(n)} e_A(n)\mathbf{x}_A(n), \quad (5)$$

$$\mathbf{w}_C(n+1) = \mathbf{w}_C(n) + \frac{\mu_C(n)}{\mathbf{x}^T(n)\mathbf{x}(n)} e_C(n)\mathbf{x}_C(n), \quad (6)$$

where the estimation error signals are defined as

$$e_B(n) = e(n) = d(n) - y_B(n), \quad (7)$$

$$e_A(n) = d(n) - y_A(n) - y_B(n) = e_B(n) - y_A(n), \quad (8)$$

$$e_C(n) = d(n) - y_C(n) - y_B(n) = e_B(n) - y_C(n), \quad (9)$$

and the input data vectors are defined as:

$$\mathbf{x}_A(n) = \mathbf{x}_C(n) = [x(n), \ldots, x(n-N_T+1)]^T, \quad (10)$$

$$\mathbf{x}_B(n) = [x(n-N_T), \ldots, x(n-N+1)]^T, \quad (11)$$

$$\mathbf{x}(n) = [\mathbf{x}_A^T(n), \mathbf{x}_B^T(n)]^T. \quad (12)$$

Equations (4), (5) derive from the NLMS algorithm applied to the filter $\left[\mathbf{w}_A^T(n), \mathbf{w}_B^T(n)\right]^T$ that processes $\mathbf{x}(n)$ and outputs $y_A(n) + y_B(n)$. Due to the N_T samples delay of the microphone signal, the optimal solution of $\mathbf{w}_A(n)$ is zero (*i.e.* $y_A(n) = 0, \forall n$). Therefore, in equation (4) $\mathbf{w}_B(n)$ can be updated with the error signal $e_B(n)$ of (7). In this way, $\mathbf{w}_A(n)$ can be adapted with any step-size $\mu_A(n)$ without affecting $\mathbf{w}_B(n)$ and its echo cancellation process. Equations (6), (9) that adapt $w_C(n)$ derive in a similar manner from filter $\left[\mathbf{w}_C^T(n), \mathbf{w}_B^T(n)\right]^T$. The adaptation rules of $\mathbf{w}_A(n)$ and $\mathbf{w}_C(n)$ differ only for the step-size choice.

The filter $\mathbf{w}_C(n)$ is employed for the double-talk detection. This filter is adapted with a fixed and high step-size ($\mu_C(n) = \mu_C \simeq 1$). During single-talk of the remote speaker, the filter $\mathbf{w}_C(n)$ tends to be zero because of the echo signal delay of at least N_T samples. During double-talk, the filter $\mathbf{w}_C(n)$ disadapts in a few sample time: its coefficients try to follow the near-end speech, which is uncorrelated with the input signal $x(n)$, and thus they get high non-zero values. We employ as index of the local speech activity the time-averaged q norm of the filter $\mathbf{w}_C(n)$ which is given by equation (13),

$$I_C(n) = < \frac{1}{N_T} \sum_{k=0}^{N_T-1} \left|w_C(n)[k]\right|^q > . \quad (13)$$

The supplementary filter $\mathbf{w}_A(n)$ is employed for estimating the system error norm and thus monitoring the echo impulse response variations. This filter is adapted with the same variable step-size of the echo-canceller, taking care that the adaptation of $\mathbf{w}_A(n)$ is never inhibited ($\mu_A(n) = \max(\mu_B(n), \mu_{A\,\min})$). The convergence state of the echo canceller is monitored with the p norm of vector $\mathbf{w}_A(n)$ as given in equation (14).

$$I_A(n) =< \frac{1}{N_T} \sum_{k=0}^{N_T-1} \left|w_A(n)[k]\right|^p > . \tag{14}$$

As for the variable step-size μ_B we have considered the expression given in equation (15),

$$\mu_B(n) = \frac{\big(\delta + A \cdot I_A(n)\big)P_x(n)}{\big(\delta + A \cdot I_A(n)\big)P_x(n) + \big(1 + C \cdot I_C(n)\big)P_e(n)}, \tag{15}$$

where δ, A and C are suitable positive constants (δ a small positive constant) and $P_x(n)$ and $P_e(n)$ are some signal power estimates of $x(n)$ and $e(n)$, respectively. For the NLMS algorithm it is convenient to estimate $P_x(n)$ as $(\mathbf{x}^T(n) \cdot \mathbf{x}(n))/N$.

The reader can easily understand the behavior of the variable step-size in equation (15). The term $I_C(n)$ at the denominator of $\mu_B(n)$ amplifies the effect of the power term $P_e(n)$. As soon as double-talk occurs, $P_e(n)$ and $I_C(n)$ assume high values that slow down the adaptation of both filters $\mathbf{w}_A(n)$ and $\mathbf{w}_B(n)$. In this situation $I_A(n)$ is kept small enough to avoid a false echo variation detection. $P_e(n)$ and $I_C(n)$ may increase also after a sudden variation of the echo impulse response; nevertheless, in this other situation $I_C(n)$ tends to zero and it allows a fast growth of the term $I_A(n)$. Therefore, the echo canceller rapidly readapts to the optimal solution.

The variable step-size of equation (15) can be interpreted as an estimate of the optimal step-size for the NLMS algorithm, reported in equation (16),

$$\mu_{Opt} = \frac{E\big\{(c(n) - y(n))^2\big\}}{E\big\{e^2(n)\big\}}. \tag{16}$$

By assuming the residual echo $\big(c(n) - y(n)\big)$ to be uncorrelated with the local disturbance $s(n)$ and $v(n)$, we obtain the expression of equation (17),

$$\mu_{Opt} = \frac{E\big\{(c(n) - y(n))^2\big\}}{E\big\{(c(n) - y(n))^2\big\} + E\big\{(s(n) + v(n))^2\big\}}. \tag{17}$$

Let us assume in equation (15) that $\delta = 0$. Then this equation can be written as follows

$$\mu_B(n) = \frac{NI_A(n)P_x(n)}{NI_A(n)P_x(n) + N\big(1 + C \cdot I_C(n)\big)P_e(n)/A}, \tag{18}$$

When $p = 2$, $N \cdot I_A(n)P_x(n)$ is the estimate of the undisturbed error signal $E\big\{(c(n) - y(n))^2\big\}$ as computed with the delay coefficient technique, while $N \cdot \big(1 + CI_C(n)\big)P_e(n)/A$ can be interpreted as an estimate of the local disturbance power $E\big\{(s(n)+v(n))^2\big\}$ (C and A have to be properly chosen for this purpose).

3. AN EXPERIMENTAL RESULT

In this Section we discuss an experimental result obtained with some real acoustic echo signal recorded in a Volkswagen Golf car. In our experiment the far-end speech, a female voice, was amplified with a hi-fi amplifier and was sent to a loudspeaker system constituted by five elements (four front and one rear loudspeaker). The microphone was positioned on the sun visor. The echo signal, the local voice and the noise were recorded separately and added together in our simulation system. In these experimental conditions the echo signal (Figure 2, plot 1) was 10 dB below the local speech (Figure 2, plot 2). A car noise was added to these signal and the signal to noise ratio was around 10 dB. The car system was modeled with a filter with length $N = 256$ and $N_T = 8$ adapted with the Decorrelation NLMS algorithm [6]. In the experiment a strong echo-path variation was simulated: after 6 seconds an echo power variation of +20 dB and a 10 tap delay of the echo signal were artificially introduced.

Figure 2, Plots 3 and 4 show the performance of the control method. The variable step-size $\mu_B(n)$ reacts rapidly to the double-talk situations and it avoids any distortion of the local speech. Nevertheless, it is still able to exploit the inter-phoneme and inter-word silence periods of the near-end speech in order to track the variations of the echo-path impulse response.

Plots 5 and 6 show the behavior of $I_A(n)$ and $I_C(n)$. It can be noticed that $I_C(n)$ effectively detects double-talk. On the contrary, $I_A(n)$ is only slightly affected by the local speech, while it reacts to the strong echo variation in about 0.5 seconds.

4. CONCLUSIONS

In this paper we presented a novel variable step-size control for acoustic echo cancellation algorithms.

The variable step-size is able to discriminate between double-talk and echo-path impulse response variations. Local speech activity and echo changes are monitored by two low memory length adaptive filters. Therefore, the overall computational complexity of the control system is very low.

The method works well with both weak and strong echoes (as large as 20dB above the level of the near-end signal) and it allows an improved adaptation by exploiting the short silence periods of the near-end speech.

The high performance and the low computational complexity make the novel variable step-size control suitable for both hands-free and hand-set cellular telephones.

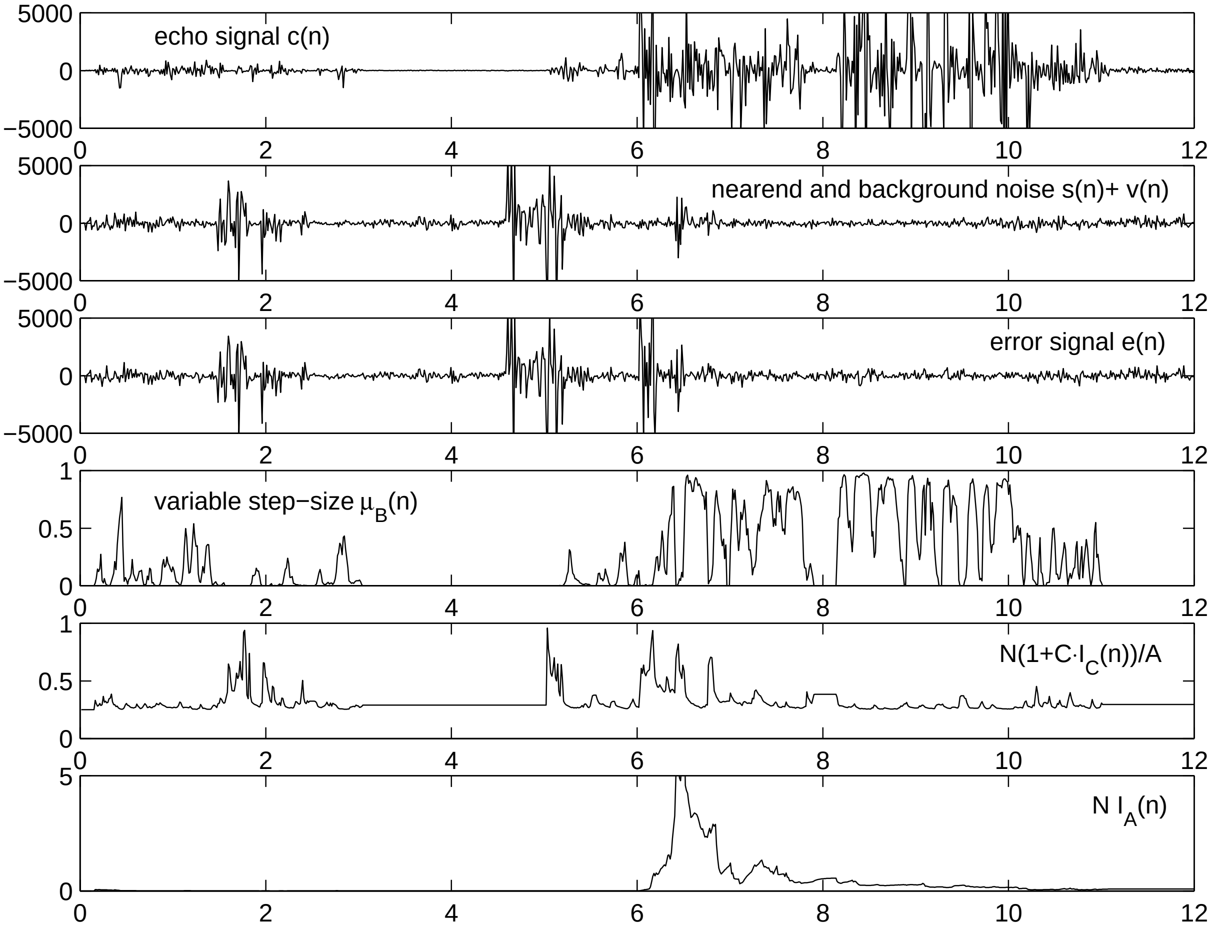

Figure 2: An experimental result: an acoustic echo with 12 second duration recorded in Volkswagen Golf car. The acoustic echo has been processed with a DLMS algorithm equipped with the novel control of adaptation. Double-talk has been introduced after 1.5, 4.5 and 6.5 seconds. An echo power variation of +20 dB and a 10 tap delay of the echo signal were introduced after 6 seconds.

5. REFERENCES

[1] A. Gilloire et alt.: *State of the art of acoustic echo cancellation.* A. R. Figueiras-Vidal (ed.), Digital Signal Processing in Telecommunications, pp. 45-91, Springer, Berlin, 1996.

[2] G. O. Glentis, K. Berberidis and S. Theodoridis: *Efficient Least squares adaptive filtering for FIR transversal filtering.* Signal Processing Magazine, vol. 16, no. 4, pp. 13-41, July 1999.

[3] C. Breining et alt.: *Acoustic echo control.* Signal Processing Magazine, vol. 16, no. 4, pp. 42-69, July 1999.

[4] A. Mader, H. Puder and G. U. Schmidt: *Step-size control for acoustic echo cancellation filters - an overview.* Signal Processing, vol. 80, pp. 1697-1719, 2000.

[5] M. M. Sondhi and D. A. Berkley: *Silencing echoes on the telephone network.* Proc. IEEE, vol. 68, no. 8, pp. 948-963, 1980.

[6] J. F. Doherty and R. Porayath: *A robust echo canceller for acoustic environments.* IEEE Trans. on Circuits and Systems II: Analog and Digital Signal Processing, vol. 44, no. 5, pp. 389-396, May 1997.

[7] P. Heitkämper: *An adaptation control for acoustic echo cancellers.* IEEE Signal Processing Letters, vol. 4, no. 6, pp. 170-172, June 1997.

[8] C. Antweiler, J. Grunwald and H. Quack: *Approximation for optimal step-size control for acoustic echo cancellation.* Proc. ICASSP-97, IEEE Conference on Acoustics Speech and Signal Processing, München, Germany, 1997, vol. 1, pp. 295-298.