A ROBUST NOISE AND ECHO CANCELLER

Khaldoon Al-Naimi, Christian Sturt and Ahmet Kondoz

Centre for Communication Systems Research (CCSR)
University of Surrey, Guildford, Surrey GU2 7XH, UK
K.Al-Naimi,C.Sturt,A.Kondoz@eim.surrey.ac.uk

10.21437/Eurospeech.2003-251

Abstract

The performance of an echo canceller systems deployed in a practical communication environment (i.e. the presence of background noise and the possible double talk scenario) depends on an accurate Voice Activity Detector (VAD) and an effective filter coefficient adaptation strategies. Accuracy of the VAD, which affects the coefficient adaptation strategy, is itself affected by the presence of background noise. In this paper, a novel soft weighting approach is proposed to replace the VAD and filter coefficient adaptation strategy. The robustness of the echo canceller system is further improved through integrating it with a noise suppression algorithm. The integrated echo canceller and noise suppressor systems has shown excellent performances under double talk scenarios with SNR as low as 5 dB.

1. Introduction

Echo is considered to be one of the most objectionable features of communication systems. It can be a result of either a mismatch at the hybrid switch (such as 4 wire to 2 wire conversion and vice versa) as in the network echo case, or the reflections caused by a reverberant environment as in acoustic echo. It manifests itself as the originator of a speech signal being able to hear his/her own speech after a certain delay. Increases in both echo delay and echo amplitude increase the level of annoyance. Background noise, as well as being subjectively objectionable, can also disrupt the proper operation of the various subsystems of the communication link, such as the speech coder. The different kinds of background noise vary widely in their characteristics, and a practical noise reduction scheme has to be capable of handling noise with different characteristics.
The paper starts with a brief introduction to the Minimum Mean Square Error Log Spectral Amplitudes (MMSE-LSA) noise suppression algorithm, as given in section 2. Section 3 gives a general description of echo canceller systems along with the problems associated with existing systems. This is then followed by our novel Adaptive Normalised Least Mean Square (ANLMS) echo canceller systems in section 4. Section 5, describes the novel integrated MMSE-LSA and ANLMS system. Experimental results for the integrated system follow in section 6. This paper ends with section 7 where conclusions are drawn.

2. MMSE-LSA Noise Suppression System

The noise suppression algorithm used in this paper is an adaptation of the Minimum Mean Square Error-Log Spectral Amplitudes (MMSE-LSA) technique first proposed by Ephraim and Malah [1][2]. The MMSE-LSA system is a member of the Short Time Spectral Amplitude (STSA) estimators that modify the spectral amplitude of the noisy speech and leave the phase untouched. The phases from the noisy speech signal are used along with the modified spectral amplitudes in reconstructing the noise suppressed speech.
Ephraim-and-Malah's MMSE-LSA noise suppression system [1][2] is given by:

$$\hat{A}_k = G_m(k) G_{lsa}(k) R_k \tag{1}$$

where R_k is the k^{th} frequency bin magnitude of the noisy input speech, $G_m(k)$ is the soft-decision gain modification and $G_{lsa}(k)$ is the log-spectral amplitude gain function [1][2].

3. Echo Canceller

Echo in telecommunication systems is the delayed and distorted sound which is reflected back to the source. Figure 1 shows a

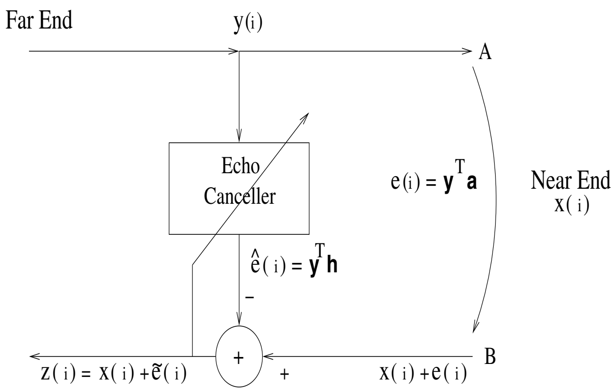


Figure 1: Block diagram of an echo canceller system

typical echo canceller configuration. The echo canceller generates the echo replica, $\hat{e}(i)$, by applying the reference signal (i.e. far end speech $y(i)$) to a transversal filter. Since the echo path between points A and B is not normally known in advance, the canceller needs to adapt the coefficients of the transversal filter. Therefore, a closed loop coefficient adaptation algorithm is required to minimise the Mean Squared Error (MSE) between the echo and its replica, i.e. $z(i)$ in Figure 1. $z(i)$ is comprised of the echo path error $\tilde{e}(i)$ and the near end speech signal $x(i)$, which is uncorrelated with the far end signal $y(i)$, and is given by

$$z(i) = x(i) + \tilde{e}(i). \tag{2}$$

Thus the MSE results in

$$E\left\{z(i)^2\right\} = E\left\{x(i)^2\right\} + E\left\{\tilde{e}(i)^2\right\} \tag{3}$$

where E denotes the expectation operator. The echo term $E\left\{\tilde{e}(i)^2\right\}$ will be minimised when $E\left\{z(i)^2\right\}$ is minimised. If we assume that there is no near end speech (i.e. $x(i) = 0$), the minimum is achieved by adjusting the coefficients h_k along the negative gradient of $E\left\{\tilde{e}(i)^2\right\}$ at each step using the equation [3]

$$h_k(i+1) = h_k(i) - 2\beta e(i) y(i-k) \tag{4}$$

where β is the step size. This is referred to as the Least Mean Square (LMS) adaptation algorithm. To improve the convergence of the filter coefficients, the normalised LMS (NLMS) algorithm has been suggested as given by

$$h_k(i+1) = h_k(i) - 2\beta \frac{\tilde{e}(i) y(i-k)}{\sigma(i)^2} \tag{5}$$

where $\sigma(i)^2$ is

$$\sigma(i)^2 = \sum_{k=0}^{N-1} y(i-k)^2 \tag{6}$$

NLMS and LMS (or variant of these) performances degrade rapidly when near end speech is present or echo is contaminated by background noise or, especially, if both cases apply. This is related to the $x(i)$ portion present in Equation (2). This uncorrelated portion results in divergence of the filter coefficients.

Stopping or reducing filter coefficient adaptation when near end speech (or double talk) is present, represent a popular solution to the divergence problem of practical echo canceller systems. This solution depends on an accurate near end detector (also known as VAD), the reliability of which is affected by background noise contamination. Another drawback of this solution is the loss of synchronisation with the echo path changes during long double talk segments which results in temporary filter coefficient divergence when adaptation is resumed.

To remove the dependence of the echo canceller performance on the reliability of the VAD and to avoid losing synchronisation with the echo path changes through continued filter coefficient adaptation we propose a novel Adaptive NLMS (ANLMS) echo canceller system.

4. Adaptive NLMS

The Adaptive NLMS (ANLMS) is based on the NLMS algorithm suggested in [3]. The NLMS of [3] differs from the general NLMS in that filter coefficients are updated less frequently with a thinning factor M, resulting in

$$h_k(i+1) = h_k(i) + \beta \frac{\sum_{m=0}^{M-1} e(i-m) y(i-m-k)}{\sigma(i)^2} \tag{7}$$

where $k = h, h+M, h+2M, ...$ and $h = 0, 1, ...M-1$.

The ANLMS includes a number of enhancements to the system of [3] (broadly: increased robustness to noise contamination, continuous filter coefficient adaptation and the removal of any need for detectors), in order to deal with the various scenarios of a practical communication system leading to

$$h_k(i+1) = h_k(i) + w_k(i)\beta \frac{\sum_{m=0}^{M-1} e(i+M-m) y(i+M-m-k)}{\psi(i)^2 \rho(i)^2} \tag{8}$$

where $\psi(i)$ and $\rho(i)$ are given by

$$\psi(i) = \alpha_e \psi(i-1) + (1-\alpha_e)\left|y(i)\right| \tag{9}$$

and

$$\rho(i) = \alpha_e \rho(i-1) + (1-\alpha_e)\left|z(i)\right| \tag{10}$$

respectively. The weighting function $w_k(i)$ is

$$w_k(i) = \exp\left\{-\left(\frac{\hat{h}_k(i) - \bar{h}_k(i)}{\gamma\beta}\right)^2\right\} \tag{11}$$

where $\hat{h}_k(i)$ is the unweighted estimate of filter coefficient k at time i and $\bar{h}_k(i)$ is the average track of filter coefficient k at time i and are respectively given by

$$\hat{h}_k(i) = h_k(i) + \beta \frac{\sum_{m=0}^{M-1} e(i+M-m) y(i+M-m-k)}{\psi^2(i)\rho^2(i)} \tag{12}$$

$$\bar{h}_k(i) = \alpha_h \bar{h}_k(i-1) + (1-\alpha_h) h_k(i). \tag{13}$$

where $0 \leq \alpha_e \leq 1$, $0 \leq \alpha_h \leq 1$ and $\gamma > 0$ are tuning parameters.

The novelty of the proposed echo-canceller method stems from the soft decision weighting function, $w_k(i)$. This weighting function removes the need for a near end speech detector and its associated problems. It also provides a soft decision means for continuous filter coefficient adaptation so as not to lose synchronisation with echo path changes. In addition it results in increased robustness to background noise contamination.

At time i, if the difference between $\hat{h}_k(i)$ and $\bar{h}_k(i)$ (as give by Equations 12 and 13 respectively) is large, which generally occurs when there is an uncorrelated part (such as near end speech, background noise or both), then $w_k(i)$ is small resulting in a small adaptive step size $w_k(i)\beta$. Whereas when the difference is small, which is mostly related to the echo, $w_k(i)$ will be closer to one and therefore increasing the adaptive step size.

Note that the novel ANLMS echo canceller system proposed in this section depends on the value of each average filter coefficient track and therefore requires an initialisation period that is dependent on the application. This initial period is essential in getting a reliable average filter coefficient track and for the overall system convergence.

5. Integrated Noise Suppressor and Echo Canceller System

The proposed ANLMS echo canceller system has no need for a VAD to control filter coefficient adaptation and its related problems. It also continues filter coefficient adaptation during double talk scenarios (presence of near end speech) and has improved robustness to background noise contamination. However, for increased robustness under heavy noise contamination while keeping near end speech unaffected, the MMSE-LSA noise suppressor is integrated into the echo canceller system as shown in Figure 2. Equation 1 of the MMSE-LSA noise suppressor algorithm has been further enhanced by introducing a weighting function resulting in:

$$\hat{A}_k = W(k) G_m(k) G_{lsa}(k) R_k \tag{14}$$

where $W(k)$ is a weighting function deduced from the shape of the echo canceller filter.

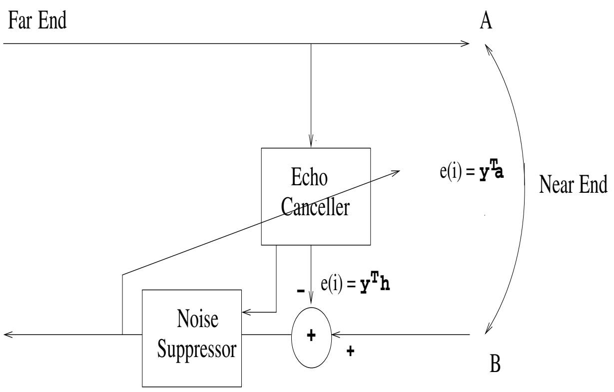
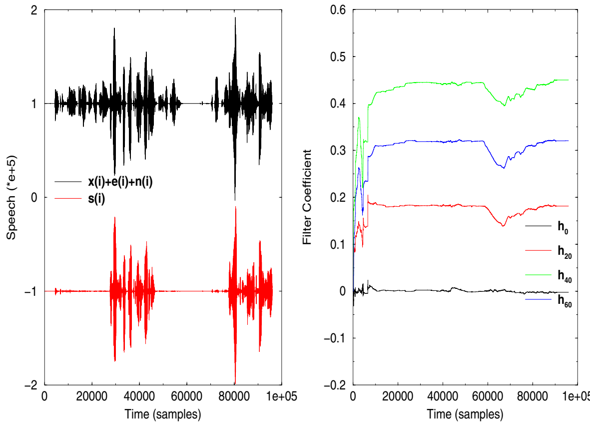


Figure 2: Block Diagram of the Novel Echo Canceller and Noise Suppressor System



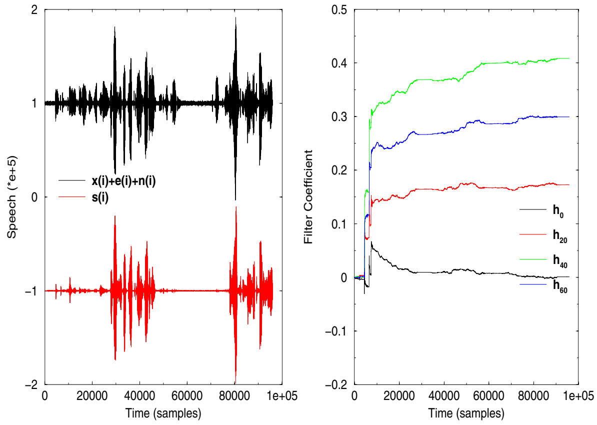
(a) (b)

Figure 3: Performance of the proposed system for clean speech (a) Output of the proposed system (b)Filter coefficient convergence tracks

6. Experimental Results

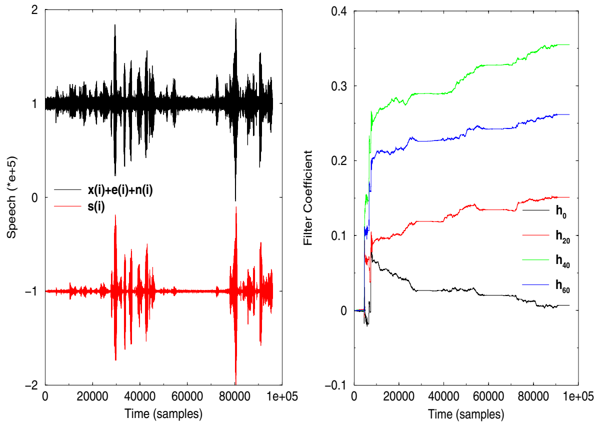
The novel integrated ANLMS and MMSE-LSA system has been tested both subjectively and objectively. For subjective testing, informal listening test were carried out. Objective testing is presented through the plots of the various filter coefficient convergence tracks along with the corresponding input (i.e. $x(i) + e(i) + n(i)$) and output (i.e. $s(i)$) of the proposed system. An echo signal resulting from the sum of three different delayed and attenuated far end speech was generated for this purpose. The echo has then been mixed with the near end speech signal along with background noise contamination resulting in SNRs of 5, 10, 20 and ∞ dB. An initialisation period of 1 sec so as to get a reliable average filter coefficient track has been assumed for all experiments.

Figures 3-6 show the results obtained from the echo signal generated through the sum of three different delays of $y(i)$: 20, 40 and 60 samples with corresponding attenuation factors of 0.2, 0.48 and 0.35 respectively. Part A of Figures 3-6 show the input to the novel system and the corresponding output, whereas part B shows the convergence track of filter coefficients h_{20}, h_{40} and h_{60} alongside an arbitrary example of another:h_0. The robustness of the system under noisy conditions as well as the convergence of the filter coefficients even in the presence of near end speech, are quite evident in Figures 3-6.



(a) (b)

Figure 4: Performance of the proposed system for 20 dB SNR (a) Output of the proposed system (b)Filter coefficient convergence tracks

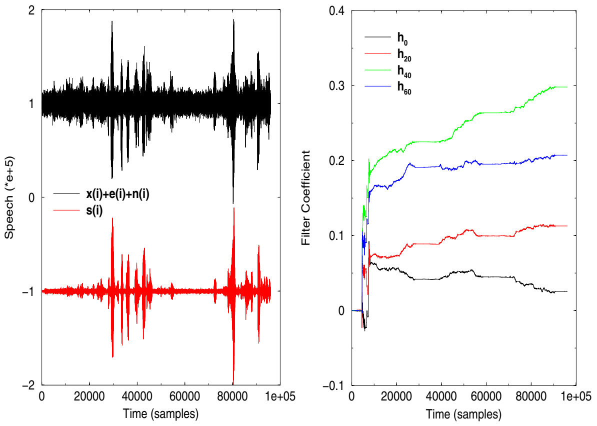


(a) (b)

Figure 5: Performance of the proposed system for 10 dB SNR (a) Output of the proposed system (b)Filter coefficient convergence tracks

7. Conclusions

We conclude, as supported by the results displayed in Figures 3-6, that the novel integrated ANLMS and MMSE-LSA system proposed is capable of continually adapting the transversal filter coefficients whether or not near speech and/or background noise are/is present. The dependence of the filter coefficient adaptation on a reliable near end VAD and its associated problem has also been eliminated. It can also be observed from Figures 3-6 that, filter coefficient convergence time increases as the amount of noise contamination increases but it is always reached. Even for the case of 5 dB SNR, the echo canceller manages to converge to match the original echo. This was also confirmed by informal listening tests of longer sequences for which convergence was always reached.



(a) (b)

Figure 6: Performance of the proposed system for 5 dB SNR (a) Output of the proposed system (b)Filter coefficient convergence tracks

8. References

[1] Yariv Ephraim and David Malah: “Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator”, IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-32, No. 6, pp 1109-1121, December 1984.

[2] Yariv Ephraim and David Malah: “Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator”, IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-33, No. 2, pp 443-445, April 1985.

[3] D. Messerschmitt, D. Hedberg, C. Cole, A. Haoui and P. Winship, Technekron Communications Systems: ”Digital Voice Echo Canceller with a TMS32020”, Application Report: SPRA129, Digital Signal Processing Solutions, Texas Instruments 1989.

[4] Regine Le Bouquin Jeannes, Pascal Scalart, Gerard Faucon and Chrisptophe Beaugeant: ”Combined Noise and Echo Reduction in Hands-Free Systems: A Survey”, IEEE Trans. on Speech and Audio Processing, Vol. 9, No. 8, November 2001.