COUPLED ADAPTIVE FILTERS FOR ACOUSTIC ECHO CONTROL AND NOISE REDUCTION

R. MARTIN and J. ALTENHÖNER

Institute of Communication Systems and Data Processing (IND), Aachen University of Technology, Templergraben 55, 52056 Aachen, Germany, Phone: +49 241 806984, Fax: +49 241 8888186, E-Mail: martin@ind.rwth-aachen.de

ABSTRACT

This contribution presents new adaptive algorithms for acoustic echo control and noise reduction which employ one, two, or possibly more microphone signals. The new algorithms accommodate high echo attenuation and lead to implementations with reduced complexity. These algorithms combine a conventional FIR echo canceller with a second NLMS-adapted FIR filter which attenuates residual echoes. The paper presents a one-microphone system with improved echo attenuation and a two-microphone system which attenuates acoustic echoes as well as ambient noise and near end speech reverberation. The algorithms can be interpreted as a frequency selective generalization of the well known voice controlled switch. This paper explains the algorithms and presents experimental results in real acoustic environments.

1. INTRODUCTION

The realization of a hands-free telephone which includes an acoustic echo cancellation and a noise reduction system still presents a major challenge. The problems are due to the (possibly) long reverberation time of the acoustic environment and the (possibly) high levels of ambient noise. Despite the recent advances in digital signal processing technology the complexity, the insufficient speed of convergence, and the lack of robustness against noise call for new solutions to the echo control problem.

In this paper we propose an approach with reduced complexity and substantially increased echo attenuation. We present two different realizations of this approach: One which uses only one microphone and another which uses two or more microphones. Besides the conventional FIR echo canceller we employ an additional NLMS-adjusted FIR filter which acts as a frequency selective echo and noise attenuator. While the echo canceller models the early reflections and the early reverberation portion of the room impulse response, the FIR attenuator reduces the residual echoes corresponding to late reverberation. The combined system achieves a high echo attenuation even in the presence of high noise levels. During 'double talk' situations, the speech quality of the FIR attenuator is superior to an echo canceller combined with conventional gain control.

The two-microphone system is based on the observation that the ambient noise as well as the late reverberation is spatially uncorrelated. Thus, the noise signals picked up by two microphones are almost uncorrelated when the microphones are sufficiently apart (e.g. 0.4 m). If the near end speaker is close to the microphones, i.e. within the reverberation distance, the near end speech signal is highly correlated on all microphone channels. The separation of ambient noise and reverberation from near end speech can then be accomplished with an adaptive (Wiener) filter. This idea was first introduced by *Allen, Berkley*, and *Blauert* [1] and was further developed e.g. by *Ferrara* and *Widrow* [2], *Zelinski* [3], [4], and *Martin* and *Vary* [5]. In this paper we will show how this approach can be effectively combined with acoustic echo compensation. In the following section we introduce the one-microphone system and the principle of frequency selective echo attenuation. In section 3 we explain the two-microphone echo and noise control algorithm. In section 4 we present our experimental results obtained in a meeting room and in a car.

2. ONE-MICROPHONE ECHO CONTROL SYSTEM

The principle of frequency selective echo attenuation is explained on the basis of the single microphone system (see Figure 1). An echo canceller with coefficient vector $\mathbf{C}(k)$

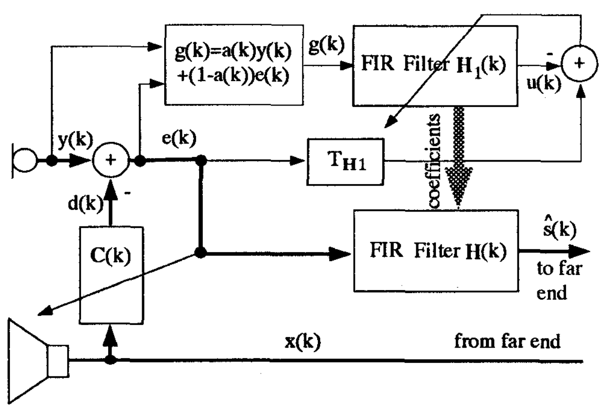

Figure 1: Echo canceller combined with FIR echo attenuator.

0-7803-2431-5/95 $4.00 © 1995 IEEE

of order N is adapted by the NLMS algorithm using linear prediction filters [6] and an adaptive step size controller to speed up convergence [7]. The coefficient update equation is given by

$$\mathbf{C}(k+1) = \mathbf{C}(k) + \alpha_{adapt}(k)\frac{e_p(k)\mathbf{X}_p(k)}{\mathbf{X}_p^T(k)\mathbf{X}_p(k)} \tag{1}$$

where $\alpha_{adapt}(k)$ denotes the adaptive step size variable. We use linear prediction filters to filter the far end speech signal $x(k)$ and the compensation error $e(k) = y(k) - \mathbf{C}^T(k)\mathbf{X}(k)$. Both prediction filters are of order two and are adapted on the far end signal $x(k)$. $\mathbf{X}_p(k)$ denotes the vector of order N of the far end prediction residual signal $x(k)$. $e_p(k)$ is the prediction residual of the compensation error $e(k)$. A second NLMS-adapted filter $\mathbf{H}_1(k)$ of order $M = 20$ minimizes the mean square error between the filter output $u(k)$ and the compensation error signal $e(k)$. The coefficient update recursion of filter $\mathbf{H}_1(k)$ is given by

$$\mathbf{H}_1(k+1) = \mathbf{H}_1(k) + \alpha_H \frac{\left(e(k-\frac{M}{2}) - \mathbf{H}_1^T(k)\mathbf{G}(k)\right)\mathbf{G}(k)}{\mathbf{G}^T(k)\mathbf{G}(k)} \tag{2}$$

where $\alpha_H = 0.1$ denotes a fixed step size parameter and $\mathbf{G}(k)$ is a vector of filter input samples $g(k)$. Due to the delay $T_{H1} = M/2$ the filter $\mathbf{H}_1(k)$ approximates a non causal Wiener filter. The coefficients of filter $\mathbf{H}_1(k)$ are copied into a second FIR filter $\mathbf{H}(k)$ which processes the compensated signal $e(k)$ prior to transmitting it to the far end speaker. The input signal $g(k)$ of filter $\mathbf{H}_1(k)$ is a mixture of compensated and not compensated signal $g(k) = a(k)y(k) + (1 - a(k))e(k)$ where $a(k)$ is an adaptive 'mixing' factor in the range [0,1]. During 'single talk' of either the near end or the far end speaker we choose $a(k) = 1$, i.e. $g(k) = y(k)$. The filter $\mathbf{H}_1(k)$ then reacts to the attenuation of the echo canceller $\mathbf{C}(k)$ such that the far end speech is attenuated and the near end signal is, at least ideally, not modified. However, to avoid distortions of low level near end speech during 'double talk' the attenuation is reduced by choosing a smaller mixing factor for double talk conditions, typically $a(k) = 0.3$. Thus, the maximum attenuation provided by the filters $\mathbf{H}_1(k)$ and $\mathbf{H}(k)$ depends on the echo attenuation of compensator $\mathbf{C}(k)$ and on the factor $a(k)$. The factor $a(k)$ is adapted using signals which measure the activity of the near end and far end speaker. A good choice for such a control signal is the adaptive step size $\alpha_{adapt}(k)$ which is piecewise linearly mapped onto the mixing factor according to

$$a(k) = \begin{cases} 0.3 & \alpha_{adapt}(k) < 0.01 \\ 0.07\alpha_{adapt}(k) + 0.3 & 0.01 \le \alpha_{adapt}(k) \le 10 \\ 1 & \alpha_{adapt}(k) > 10 \end{cases} \tag{3}$$

The adaptive filter $\mathbf{H}(k)$ can be interpreted as a frequency selective generalization of a voice controlled switch whereas the NLMS algorithm in connection with this simple signal mixing mechanism was found to be an extremely well suited adaptation algorithm. Under simplifying assumptions it can be shown that for $a(k) = 1$ the filter $\mathbf{H}(k)$ almost doubles the compensator attenuation.

3. THE TWO-MICROPHONE SYSTEM

The two-microphone system allows to take full advantage of the proposed concept since it also reduces noise and reverberation. The two-microphone system is a combination of a two-microphone noise reduction system [5], [8] and the principle of frequency selective echo attenuation as explained in the previous section. The microphone signals $y_1(k)$ and $y_2(k)$ are compensated using compensators $\mathbf{C}_1(k)$ and $\mathbf{C}_2(k)$. The microphone distance is chosen such that ambient noise components within the microphone signals are mutually uncorrelated. If the near end speaker is

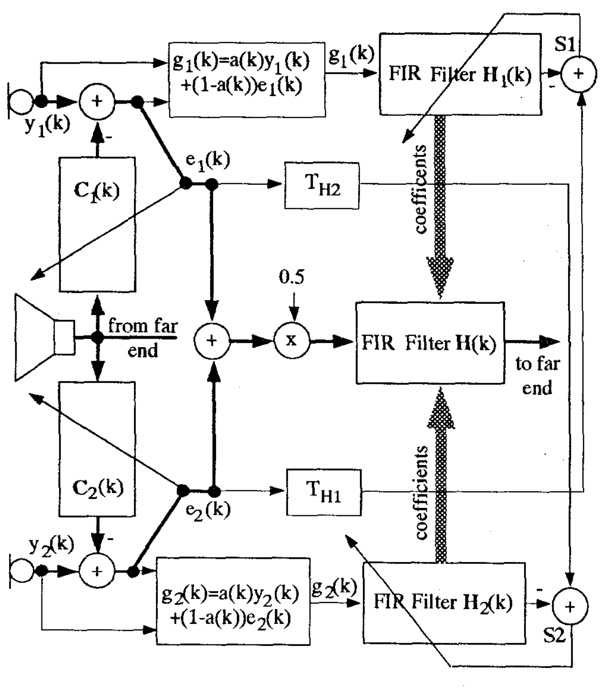

Figure 2: Basic block diagram of the two-microphone coupled filter approach.

close to the microphones, i.e. if his/her distance to the microphones is less than the reverberation distance, the near end speech signal is highly correlated on both microphone channels and can be separated from the uncorrelated noise components by means of a Wiener filter [5]. In our implementation the Wiener filter is approximated by the FIR filter $\mathbf{H}(k)$ whose coefficients are computed as the mean of two NLMS-adapted filters $\mathbf{H}_1(k)$ and $\mathbf{H}_2(k)$, i.e.

$\mathbf{H}(k) = 0.5(\mathbf{H}_1(k) + \mathbf{H}_2(k))$. These filters are adapted using a linear-phase version of the NLMS and 65 coefficients.

Similar to the single microphone approach of section 2 the two FIR filters $\mathbf{H}_1(k)$ and $\mathbf{H}_2(k)$ will also attenuate far end echoes. The major difference with respect to the single microphone approach is, that these filters now use the other microphone channel as their reference signal. Since ambient noise and late reverberation are uncorrelated the filters $\mathbf{H}_1(k)$ and $\mathbf{H}_2(k)$ can reduce the mean square error at the summation points S1 and S2 only if they suppress far end echoes and uncorrelated ambient noise and near end speech reverberation. Near end direct sound is highly correlated on the microphone channels and thus passes these filters without modification. If the near end speaker is not in a symmetric position with respect to the microphones a time delay compensation becomes necessary. The time delay compensation was implemented by means of an SNR-sensitive cross-correlation estimator [9], [10]. For the sake of simplicity the time delay compensation in not shown in Figure 2. Since the proposed concept employs two compensators its value depends to a large extent on the compensator order needed to achieve sufficient echo attenuation.

4. EXPERIMENTAL RESULTS

We now summarize the experimental results for the one and the two microphone system. We compare the 'conventional' echo canceller with the one and the two microphone combined system in terms of Echo Return Loss Enhancement (ERLE) averaged over a speech data base of 16 phonetically balanced sentences. All speech samples are 4s long. They contain no significant speech pauses. The speech data base was recorded in a meeting room (35 m^2) with a reverberation time of $T_H = 0.7s$. The microphone signals have a far end echo to noise ratio of about 30 dB. In all experiments the echo canceller was initialized with a perfect sequence (see [11] for a definition of perfect sequences and their relation to system identification) to guarantee optimal initial convergence. We will denote the ERLE of the echo canceller and of the combined systems with $ERLE\_C(k)$ and $ERLE\_CH(k)$, respectively.

Figure 3 shows ERLE vs. time plots for the one-microphone system, single talk conditions, and various values of the mixing factor $a(k)$. An echo canceller $\mathbf{C}(k)$ of order $N = 2800$ was used in this experiment. While the echo canceller alone gives no more than 30 dB ERLE the one microphone system achieves the maximum attenuation of more than 40 dB for $a(k) = 1$. The adaptive control mechanism reaches almost the maximum attenuation during single talk conditions. Figure 4 gives time averaged ERLE values for different numbers of compensator taps. The time averaged ERLE values $\overline{ERLE\_C(k)}$ and $\overline{ERLE\_CH(k)}$ are computed as an average over the last 3 seconds of the 4 s speech samples. As it can be seen from Figure 4 and the discussion in section 2 the performance of the coupled filter algorithm depends on the compensator attenuation. To

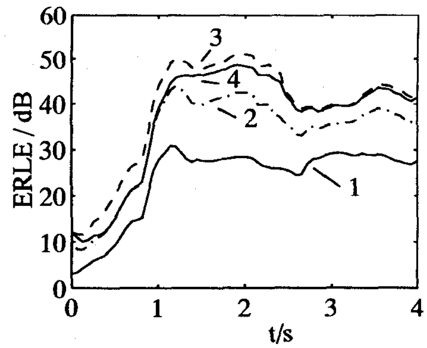

Figure 3: Echo return loss enhancement (ERLE) of the one microphone system in a meeting room: 1: $ERLE\_C(k)$ (echo canceller only); 2, 3, 4: $ERLE\_CH(k)$ (combined system) with $a(k) = 0.5$ (2), with $a(k) = 1$ (3), and with $a(k)$ adaptive (4).

achieve a significant gain the compensator $\mathbf{C}(k)$ should provide a minimum echo attenuation of 10–15 dB. To achieve an attenuation of 30 dB the compensator $\mathbf{C}(k)$ alone would need about 3000 filter taps. With the combined system, however, a compensator of order $N < 2000$ is sufficient. The distortions of near end speech during 'near end single

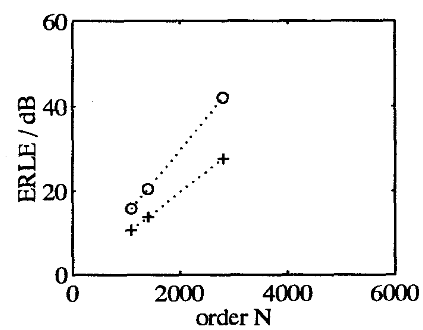

Figure 4: $\overline{ERLE\_C(k)}$ and $\overline{ERLE\_CH(k)}$ as a function of compensator order N; $a(k)$ adaptive.

talk' are negligible. During 'double talk' the contribution of filter $\mathbf{H}(k)$ to the far end echo attenuation is reduced to 3–6 dB. Far end echo attenuation can then be achieved only in frequency bands in between the formant frequencies of the near end speech signal. The distortion of the near end speech signal is acceptable.

We now turn to the two-microphone system. Figure 5 plots the time averages $\overline{ERLE\_C(k)}$ and $\overline{ERLE\_CH(k)}$ as a function of compensator order N and mixing factor $a(k)$. For $a(k) = 0$ the filter $\mathbf{H}(k)$ attenuates the spatially uncorrelated late reverberation and noise components within the microphone signals and achieves thus an attenuation of about 5–6 dB [8]. In case of an adaptively controlled mixing

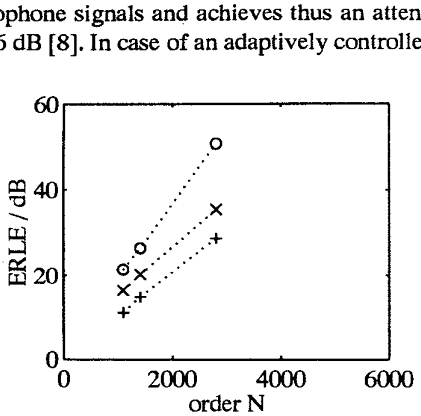

Figure 5: ERLE vs . compensator order of the two-microphone system for an meeting room: (+) $\overline{ERLE\_C(k)}$ (Echo canceller only); $\overline{ERLE\_CH(k)}$ for $\alpha(k) = 0$ (x) and for $a(k)$ adaptive (o).

factor the filter $\mathbf{H}(k)$ virtually doubles the compensator attenuation. Finally, Figure 6 gives time averaged ERLE results for signals recorded in a car. The reverberation time in this car is $T_H = 0.068$. The noise signals were

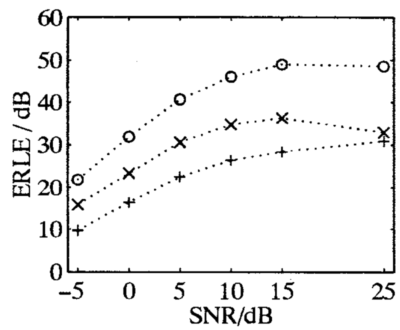

Figure 6: Echo return loss enhancement (ERLE) of two-microphone system in a car: (+): $\overline{ERLE\_C(k)}$ (echo canceller only); (x): $\overline{ERLE\_CH(k)}$ (combined system) with $a(k) = 0$; (o): $\overline{ERLE\_CH(k)}$ (combined system) with $a(k)$ adaptive.

recorded separately from the speech signals and added to the speech signals at various SNR values. The echo cancellers performance degrades with lower SNR values but most of the residual echoes are masked by the car noise. At the output of the two-microphone systems the noise level is attenuated by about 15 dB and no residual echoes are audible. Due to the underlying principle the noise reduction is most effective for microphone distances of 0.4–0.5 m and noise sources with almost flat noise power spectra. In our car environment, however, low frequency noise is predominant. As a result, the residual noise has audible 'musical' fluctuations.

5. CONCLUSIONS

The utility of the proposed approach is documented by the high echo attenuation and the robustness against noise. Regardless of compensator order it seems worthwhile to divert 21 coefficients of the compensator to the frequency selective echo attenuator. In contrast to standard systems which use an echo canceller and a (shallow) voice switch the new algorithms give better speech quality during 'double talk' situations. The two-microphone approach also reduces noise and near end speaker reverberation and can be easily extended to more than two microphones and to frequency domain implementations.

REFERENCES

[1] J. Allen, D. Berkley, and J. Blauert: "Multimicrophone signal-processing technique to remove room reverberation from speech signals", J. Acoust. Soc. Am., Vol. 62, No. 4, October 1977.

[2] E. Ferrara and B. Widrow: "Multichannel Adaptive Filtering for Signal Enhancement", IEEE Trans. Acoustics, Speech, Signal Processing, Vol. ASSP-29, no. 3, pp.766-770, June 1981.

[3] R. Zelinski: "A Microphone Array with Adaptive Post-Filtering for Noise Reduction in Reverberant Rooms", Proc. ICASSP-88, pp. 2578-2581.

[4] R. Zelinski: "Noise Reduction Based on Microphone Array with LMS Adaptive Post-Filtering", Elect. Lett., Vol. 26, No.24, pp. 2036-2037, November 1990.

[5] R. Martin and P. Vary: "A Symmetric Two Microphone Speech Enhancement System — Theoretical Limits and Application in a Car Environment", Proc. Fifth IEEE Signal Processing Workshop, Illinois, September 13-16, 1992.

[6] C. Acker (Antweiler) and P. Vary: "Combined Implementation of Predictive Speech Coding and Acoustic Echo Cancellation", Signal Processing VI, J. Vandewalle et al. (eds.), pp. 1641-1644, Elsevier 1992.

[7] S. Yamamoto and S. Kitayama: "An adaptive echo canceller with variable step gain method", Trans. IECE Japan, vol. E65, pp. 1-8, 1982.

[8] R. Martin and P. Vary: "Combined Acoustic Echo Cancellation, Dereverberation, and Noise Reduction: a Two Microphone Approach", Annales des Télécommunications, Vol. 49, no. 7-8, pp. 429-438, 1994.

[9] G. Carter: "Coherence and Time Delay Estimation", Proc. IEEE, Vol. 75, No. 2, pp. 236-255, February 1987.

[10] R. Martin: "An Efficient Algorithm to Estimate the Instantaneous SNR of Speech Signals", Proc. EUROSPEECH '93, pp. 1093-1096, Berlin, September 21-23, 1993.

[11] C. Antweiler and M. Dörbecker: "Perfect Sequence Excitation of the NLMS Algorithm and its Application to Acoustic Echo Control", Annales des Télécommunication, Vol. 49, no. 7-8, July-August 1994.