NEW ASPECTS OF COMBINING ECHO CANCELLERS WITH BEAMFORMERS

Karl-Dirk Kammeyer,

University of Bremen
Dept. of Communications Engineering
28359 Bremen, Germany

Markus Kallinger, and Alfred Mertins

University of Oldenburg
Institute of Physics, Signal Processing Group
26111 Oldenburg, Germany

ABSTRACT

In this contribution, we introduce a novel framework for combining approaches for acoustic echo cancellation and beamforming. Classical schemes of combination incorporate the simple concatenation of both subsystems. However, if the echo cancellers come first, they cannot exploit the noise reduction capabilities of the beamformer. In the other setup, a time-variant beamformer can heavily disturb the succeeding echo canceller's convergence. The new approach establishes a possibility to smoothly switch between two possible target functions: the mean squared errors at the beamformer's inputs and its output, respectively. Thus, one combined system benefits from advantages of both classical schemes. The paper contains a theoretical analysis of possible solutions for preceding echo cancellers, which are adapted using the beamformer output. The appropriate modification of the normalized least mean squares algorithm is derivated, too.

1. INTRODUCTION

Powerful systems for acoustic echo control usually employ several subsystems. The most basic subsystem is the acoustic echo canceller (AEC) [1], which represents the optimal solution in the sense of system theory. Most recent solutions for hands-free telephony employ microphone arrays. Besides reducing ambient noise, beamforming microphone arrays provide a certain amount of additional echo attenuation. Achieving a synergetic performance should constitute the goal of a combined system with echo cancellers and a beamformer. A survey of systems including both subsystems is given in [2]. Two basic systems are depicted in Figs. 1 and 2 containing configurations with a multi-microphone AEC and the beamformer just behind the microphones, respectively. M is the number of microphones and the error signals are defined by $E_0(e^{j\Omega})$ to $E_{M-1}(e^{j\Omega})$, respectively. Note, that a near-end speaker $S_n(e^{j\Omega})$ is not considered in this contribution.

A promising approach based on the beamformer-first-setup (Fig. 2) aims at the interpretation of the AEC as an

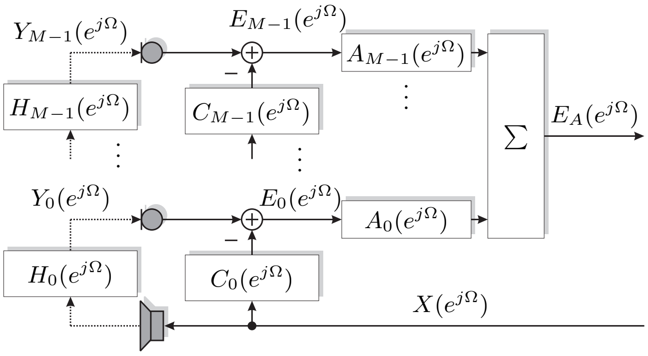

Fig. 1. Block diagram of the AEC-first-setup.

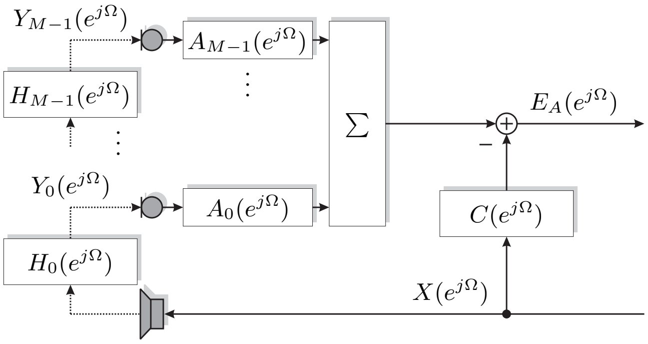

Fig. 2. Block diagram of the beamformer-first-setup.

additional microphone channel of an adaptive beamformer [3]. Here, one common error signal is used for the adaptation of both the beamformer and the AEC. This measure reveals advantages in very noisy environments. However, the approach was found to be effective, when the lengths of the beamformer and the AEC filters are equal [4]. If the order of the AEC exceeds the one of the beamformer, its convergence is heavily disturbed by the time variant beamformer. Thus, this combined system seems to be appropriate for less reverberant environments.

This paper presents a new framework which shows that adaptation according to the beamformer output $E_A(e^{j\Omega})$ is even possible using the AEC-first-setup (Fig. 1). In addition, we show that, after converging according to $E_A(e^{j\Omega})$, switching the adaptation rule using $E_0(e^{j\Omega})$, $E_1(e^{j\Omega})$, ..., $E_{M-1}(e^{j\Omega})$ is possible without harming the minimum mean squared error (MMSE)-criterion for $E_A(e^{j\Omega})$. This procedure seems advantageous, because initial adaptation can be carried out according to $E_A(e^{j\Omega})$, as long as the ambient noise level is high. Once sufficient echo attenuation is achieved, the echo cancellers can slowly converge using $E_0(e^{j\Omega})$ to $E_{M-1}(e^{j\Omega})$ in the background without affecting the overall echo attenuation.

In the next section, we examine theoretical MMSE-solutions for the AEC-first-setup, where $\mathrm{E}\left\{|E_A(e^{j\Omega})|^2\right\}$ is minimized with respect to the AEC-filters. In Section 3, we introduce a modification of the normalized least mean squares (NLMS) algorithm in order to achieve the newly introduced MMSE-solutions of Section 2. Simulation results are presented in Section 4, and Section 5 concludes the paper.

2. GENERALIZED ADAPTATION OF ECHO CANCELLERS IN FRONT OF A BEAMFORMER

In this section, we investigate a combined system for acoustic echo control employing M echo cancellers, one for each microphone, in front of a beamformer. Echo cancellers in such systems usually aim at the minimization of the mean squared error (MSE) at the output of the echo cancellers $C_0(e^{j\Omega})$ to $C_{M-1}(e^{j\Omega})$. The discrete-time Fourier transform (DTFT) of each room impulse response (RIR) is expressed by $H_0(e^{j\Omega})$ to $H_{M-1}(e^{j\Omega})$; $A_0(e^{j\Omega})$ to $A_{M-1}(e^{j\Omega})$ are the beamformer's transfer functions. For further derivations we define the following vectors

$$\begin{aligned}
\mathbf{H}(e^{j\Omega}) &= \left[H_0(e^{j\Omega}), \ldots, H_{M-1}(e^{j\Omega})\right]^T, && (1)\\
\mathbf{C}(e^{j\Omega}) &= \left[C_0(e^{j\Omega}), \ldots, C_{M-1}(e^{j\Omega})\right]^T, && (2)\\
\mathbf{A}(e^{j\Omega}) &= \left[A_0(e^{j\Omega}), \ldots, A_{M-1}(e^{j\Omega})\right]^T, \text{ and} && (3)\\
\mathbf{E}(e^{j\Omega}) &= \left[E_0(e^{j\Omega}), \ldots, E_{M-1}(e^{j\Omega})\right]^T. && (4)
\end{aligned}$$

The array's output can be expressed by

$$E_A(e^{j\Omega}) = \mathbf{A}^T(e^{j\Omega})\left(\mathbf{H}(e^{j\Omega}) - \mathbf{C}(e^{j\Omega})\right)X(e^{j\Omega}). \qquad (5)$$

Notation: ^{T}, ^{H}, and * denote transposition, Hermitian transposition, and complex conjugation, respectively. To improve the clarity of the presentation, we will omit the argument $(e^{j\Omega})$ in the following.

As with the beamformer-first-setup illustrated in Fig. 2, E_A is the actual error signal at the system output, which is transmitted back to the far-end-speaker. Therefore, we will examine the adaptation of the echo cancellers C_0 to C_{M-1} with respect to the final error signal E_A.

$$\begin{aligned}
\mathrm{E}\left\{|E_A|^2\right\} &= \mathbf{H}^H\mathbf{A}^*\Phi_{XX}\mathbf{A}^T\mathbf{H} - \mathbf{H}^H\mathbf{A}^*\Phi_{XX}\mathbf{A}^T\mathbf{C}\\
&\quad -\mathbf{H}^T\mathbf{A}\Phi_{XX}\mathbf{A}^H\mathbf{C}^*\\
&\quad +\mathbf{C}^H\mathbf{A}^*\Phi_{XX}\mathbf{A}^T\mathbf{C} \text{ with} && (6)\\
\Phi_{XX} &= \mathrm{E}\left\{|X|^2\right\}. && (7)
\end{aligned}$$

A partial differentiation with respect to $\mathbf{C}$ leads to

$$\frac{\partial \mathrm{E}\left\{|E_A|^2\right\}}{\partial \mathbf{C}} = \left(-2\mathbf{A}^*\mathbf{A}^T\mathbf{H} + 2\mathbf{A}^*\mathbf{A}^T\mathbf{C}\right)\Phi_{XX} \text{ using} \qquad (8)$$

$$\frac{\partial C}{\partial C} = 0 \quad \text{and} \quad \frac{\partial C^*}{\partial C} = 2. \qquad (9)$$

The MMSE-solution $\mathbf{C}_{E_A}$ of the AEC-first-setup according to the succeeding beamformer's output signal E_A results in

$$\mathbf{A}^*\mathbf{A}^T\mathbf{C}_{E_A} = \mathbf{A}^*\mathbf{A}^T\mathbf{H}. \qquad (10)$$

Note that the square matrix $(\mathbf{A}^*\mathbf{A}^T)$ is a dyadic product of two vectors. Thus, it has a rank of one. $\mathbf{C}_{E_A}$ can be chosen from an $(M-1)$-dimensional solution space.

In the following paragraphs we discuss three of numerous possible side conditions in order to find a unique solution for $\mathbf{C}_{E_A}$.

1. We can formulate a solution with the lowest energy in the coefficient vector $\mathbf{C}_{E_A}$

$$\begin{aligned}
\mathbf{C}_{E_A|\mathrm{LE}} &= \left(\mathbf{A}^*\mathbf{A}^T\right)^+\mathbf{A}^*\mathbf{A}^T\mathbf{H} && (11)\\
&= \frac{1}{\mathbf{A}^T\mathbf{A}^*}\mathbf{A}^*\mathbf{A}^T\mathbf{H}. && (12)
\end{aligned}$$

   The simplification of the Moore-Penrose-Pseudoinverse denoted by $^+$ in equation (12) becomes possible because we have the dyadic product of two vectors. This special solution could become interesting when large arrays come into operation at a limited quantization depth.

2. For the sake of low computational complexity, we can restrict the adaptation to only one echo canceller, e. g. C_0. The remaining echo cancellers are kept at zero, i. e. they are switched off:

$$\begin{aligned}
\mathbf{C}_{E_A|1\mathrm{EC}} &= \left[C_{0|1\mathrm{EC}}, 0, \ldots, 0\right]^T, && (13)\\
C_{0|1\mathrm{EC}} &= \frac{1}{A_0}\mathbf{A}^T\mathbf{H}. && (14)
\end{aligned}$$

   A common minimum variance distortionless response (MVDR) beamformer design does not provide zeros of A_0's z-transform on the unit circle. Otherwise, certain frequencies would not be picked up by the first microphone channel. Generally, a beamformer filter is a mixed phase system. Accordingly, the time domain representation of $C_{0|1\mathrm{EC}}$ might not have a causal shape like an ordinary RIR.

3. It seems quite obvious that

$$\mathbf{C}_{E_A|\mathbf{H}} = \mathbf{H} \tag{15}$$

has to be within the $(M-1)$-dimensional solution space, too. Note that it is the only solution that is independent of the beamformer coefficients $\mathbf{A}$.

Another interesting aspect can be retrieved from a different representation of

$$\mathbf{C}_{E_A} \;=\; \mathbf{H} + \sum_{i=0}^{M-2} \kappa_i \mathbf{K}_i \text{ with} \tag{16}$$

$$\mathbf{A}^T \mathbf{K}_i = 0 \quad \forall i. \tag{17}$$

$\mathbf{K}_i$ are the othogonal basis vectors of the $(M-1)$-dimensional solution space; κ_i are arbitrary factors. Thus, two different solutions in the solution space can be represented by two different sets of κ_i, denoted by $\kappa_{i,0}$ and $\kappa_{i,1}$. A gradient between arbitrary points in the solution space remains within the solution space

$$\mathbf{C}_{E_A,0} - \mathbf{C}_{E_A,1} = \sum_{i=0}^{M-2} \left(\kappa_{i,0} - \kappa_{i,1}\right) \mathbf{K}_i. \tag{18}$$

Consequently, any steepest descent method [5], which leads from $\mathbf{C}_{E_A,1}$ to $\mathbf{C}_{E_A,0}$, would not leave the $(M-1)$-dimensional solution space. Hence, it is possible to switch between the introduced side conditions and achieve different solutions of equation (10) without raising the MSE $\mathrm{E}\left\{|E_A|^2\right\}$.

3. MODIFICATION OF THE NLMS ALGORITHM

In this section, we describe a modification of the NLMS algorithm in order to minimize the MSE at the output of the beamformer as shown in Fig. 1. At this point, we have to switch to a discrete time domain representation with a discrete time index k. Except for now using lower case letters for signals and systems, the setup in Fig. 1 remains valid. Consequently, the output of the beamformer results in

$$e_A(k) = \sum_{i=0}^{M-1} \mathbf{a}_i^T \mathbf{X}(k) \left(\mathbf{h}_i - \mathbf{c}_i(k)\right) \tag{19}$$

with the vectors

$$\mathbf{h}_i \;=\; \left[h_i(0), \ldots, h_i(L_h-1)\right]^T, \tag{20}$$

$$\mathbf{c}_i(k) \;=\; \left[c_{i,0}(k), \ldots, c_{i,L_h-1}(k)\right]^T, \tag{21}$$

$$\mathbf{a}_i \;=\; \left[a_i(0), \ldots, a_i(L_a-1)\right]^T, \tag{22}$$

$$\mathbf{X}(k) \;=\; \left[\mathbf{x}(k), \ldots, \mathbf{x}(k-L_a+1)\right]^T, \text{ and} \tag{23}$$

$$\mathbf{x}(k) \;=\; \left[x(k), \ldots, x(k-L_h+1)\right]^T. \tag{24}$$

The modified gradient can be expressed by

$$\begin{bmatrix} \dfrac{\partial \mathrm{E}\left\{e_A^2(k)\right\}}{\partial \mathbf{c}_0(k)} \\ \vdots \\ \dfrac{\partial \mathrm{E}\left\{u^2(k)\right\}}{\partial \mathbf{c}_{M-1}(k)} \end{bmatrix} = \begin{bmatrix} \mathrm{E}\left\{-2e_A(k)\mathbf{X}^T(k)\mathbf{a}_0\right\} \\ \vdots \\ \mathrm{E}\left\{-2e_A(k)\mathbf{X}^T(k)\mathbf{a}_{M-1}\right\} \end{bmatrix}. \tag{25}$$

Finally, a modified stochastic gradient algorithm takes on the form

$$\begin{bmatrix} \mathbf{c}_0(k+1) \\ \vdots \\ \mathbf{c}_{M-1}(k+1) \end{bmatrix} = \begin{bmatrix} \mathbf{c}_0(k) \\ \vdots \\ \mathbf{c}_{M-1}(k) \end{bmatrix} + \begin{bmatrix} \mu_0 \mathrm{E}\left\{e_A(k)\mathbf{X}^T(k)\mathbf{a}_0\right\} \\ \vdots \\ \mu_{M-1} \mathrm{E}\left\{e_A(k)\mathbf{X}^T(k)\mathbf{a}_{M-1}\right\} \end{bmatrix} \tag{26}$$

where μ_0 to μ_{M-1} represent the step-size factors. Note that the convolution of $\mathbf{x}(k)$ with $\mathbf{a}_i$ introduces additional temporal correlation into the actual reference signal, which is fed into the AEC filter. Therefore, the step-size parameter has to be reduced compared to the original one, which is defined by a scalar factor $0 \leq \alpha \leq 1$ divided by the energy of the signal vector $\mathbf{x}(k)$ [5]. Finally, in order to obtain the modified NLMS algorithm, simply the expectation operators $\mathrm{E}\{\cdot\}$ have to be omitted.

Later on in Section 4, we will show simulation results converging towards the impulse responses of two solutions obtained with different side conditions, as introduced in Section 2: a solution with only one adaptive filter $\mathbf{C}_{E_A|1\mathrm{EC}}$ and the RIR identification solution $\mathbf{C}_{E_A|\mathbf{H}}$. The corresponding NLMS update rule using only one adaptive filter is

$$\mathbf{c}_0(k+1) = \mathbf{c}_0(k) + \mu_0 e_A(k)\mathbf{X}^T(k)\mathbf{a}_0. \tag{27}$$

All other filters $\mathbf{c}_1(k)$ to $\mathbf{c}_{M-1}(k)$ are kept zero. The update rule for the RIR identification solution then becomes

$$\begin{bmatrix} \mathbf{c}_0(k+1) \\ \vdots \\ \mathbf{c}_{M-1}(k+1) \end{bmatrix} = \begin{bmatrix} \mathbf{c}_0(k) \\ \vdots \\ \mathbf{c}_{M-1}(k) \end{bmatrix} + \mu \begin{bmatrix} e_0(k)\mathbf{x}(k) \\ \vdots \\ e_{M-1}(k)\mathbf{x}(k) \end{bmatrix}. \tag{28}$$

4. SIMULATION RESULTS

In this section we observe the echo return loss enhancement (ERLE) as a function of time. We introduce the ERLE at three points of interest as follows

$$\mathrm{ERLE}_{\mathrm{AEC}}(k) = \frac{\mathrm{E}\{y_0^2(k)\}}{\mathrm{E}\{e_0^2(k)\}}, \tag{29}$$

$$\mathrm{ERLE}_{\mathrm{Beam}}(k) = \frac{\mathrm{E}\{e_0^2(k)\}}{\mathrm{E}\{e_A^2(k)\}}, \text{ and} \tag{30}$$

$$\mathrm{ERLE}_{\mathrm{Sys}}(k) = \frac{\mathrm{E}\{y_0^2(k)\}}{\mathrm{E}\{e_A^2(k)\}}. \tag{31}$$

Fig. 3 shows the three ERLE measures as functions of time. We used simulated impulse responses [6] at a reverberation time of $\tau_{60} = 400\,\mathrm{ms}$. Each echo canceller contained 1024 coefficients, and $M = 2$ microphones in endfire steering came into operation. We employed a superdirective design assuming an uncorrelated noise with a power of -30 dB [7]. Here, no near-end speaker was considered, since double talk detection (DTD) schemes can easily be incorporated into the new approach by modifying μ_0. The minimization criterion was switched at 60,000 samples. First, only one echo canceller was adapted according to $e_A(k)$, and after switching, $e_0(k)$ and $e_1(k)$ were used. White Gaussian noise was used for the reference signal $x(k)$.

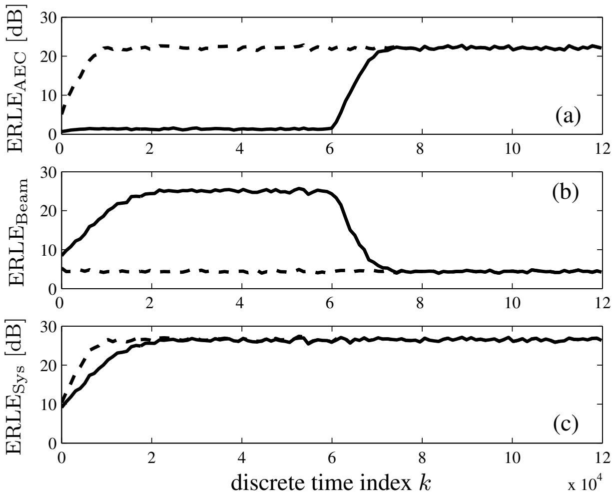

Fig. 3. Echo return loss enhancement as a function of time at the AEC (plot (a)), at the beamformer (plot (b)), and of the combined system (plot (c)). The dashed line illustrates perpetual adaptation according to $e_0(k)$ and $e_1(k)$.

The most important observation is that $\mathrm{ERLE}_{\mathrm{Sys}}(k)$ does not expose any gaps, while the criterion is being switched. One can also see that adaptation speed is slightly decreased when the modified NLMS algorithm comes into operation.

In Fig. 4 a recorded speech signal is used for the reference signal $x(k)$. Again, switching from the modified NLMS algorithm to conventional adaptation according to $e_0(k)$ and $e_1(k)$ does not exhibit any gaps.

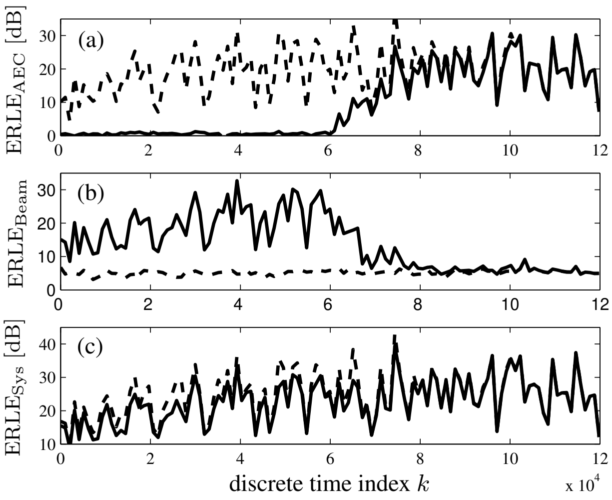

Fig. 4. Echo return loss enhancement as a function of time as in Fig. 3. A speech signal was used to feed the reference signal $x(k)$.

5. CONCLUSIONS

In this contribution we have given a new framework for the combination of acoustic echo cancellers and a beamformer. The novel adaptation scheme can be incorporated into recently proposed approaches for joint echo attenuation and noise control. It can smoothly switch between the AEC-first and the beamformer-first setup, and can efficiently exploit the benefits of both schemes.

6. REFERENCES

[1] M.M. Sondhi, "An Adaptive Echo Canceller," *Bell Syst. Tech. J*, vol. 46, no. 3, pp. 497–511, 1967.

[2] W. L. Kellermann, "Acoustic Echo Cancellation for Beamforming Microphone Arrays," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. S. Brandstein and D. Ward, Eds., chapter 13, pp. 281–306. Springer-Verlag, 2001.

[3] S. Doclo, M. Moonen, and E. De Clippel, "Combined Acoustic Echo and Noise Reduction using GSVD-based Optimal Filtering," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Istanbul, Turkey, June 2000, pp. 1051–1054.

[4] W. Herbordt, W. Kellermann, and S. Nakamura, "Joint Optimization of LCMV Beamforming and Acoustic Echo Cancellation," in *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, Vienna, Austria, September 2004, pp. 2003–2006.

[5] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 2002.

[6] J. B. Allen and D. A. Berkley, "Image Method for Efficiently Simulating Small–Room Acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, 1979.

[7] J. Bitzer and K. U. Simmer, "Superdirective microphone arrays," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. S. Brandstein and D. Ward, Eds., chapter 2, pp. 19–38. Springer-Verlag, 2001.