A frequency domain stereophonic acoustic echo canceler exploiting the coherence between the channels

Jacob Benesty

Bell Laboratories, Murray Hill, New Jersey 07974
jbenesty@bell-labs.com

André Gilloire

France Telecom CNET, France
andre.gilloire@cnet.francetelecom.fr

Yves Grenier

ENST/SIG, France
grenier@sig.enst.fr

Abstract: Stereophonic acoustic echo cancellation (AEC) is typically intended for use in high quality teleconferencing systems and multi-participant desktop conferencing that implement sound transmission through two channels. Because of the high correlation between the two channel signals, rapidly converging adaptive filter algorithms (such as the two-channel fast recursive least-squares - FRLS) are required. In this paper, we propose a new and efficient frequency domain adaptive algorithm that achieves this goal.

PACS numbers: 43.60.Gk, 43.60.Lq

1. Introduction

Stereophonic AEC can be viewed as a straightforward generalization of the single-channel acoustic echo cancellation principle.[1] Figure 1 shows this technique for one microphone in the *receiving* room (which is represented by the two echo paths, h_1 and h_2, between the two loudspeakers and the microphone). The two reference signals, x_1 and x_2, from the transmission room are obtained by two microphones in the case of teleconferencing. These signals are derived by filtering from a common source, which gives rise to a non-uniqueness problem that does not arise for the single-channel AEC.[1,2] As a result, conventional two-channel least mean square (LMS) type adaptive algorithms converge very slowly to the solution, and the two-channel FRLS algorithm (which is very complex and unstable) must be used. This requirement implies a high level of computational complexity, so that a real-time implementation of this algorithm is difficult.

It is important to devise an adaptive algorithm that takes into account the high correlation between the two channel signals or, equivalently, the coherence in the frequency domain to speed up its convergence rate. In this paper, we propose a new frequency domain adaptive algorithm that converges to a low level misalignment and is much less complex than the two-channel FRLS.

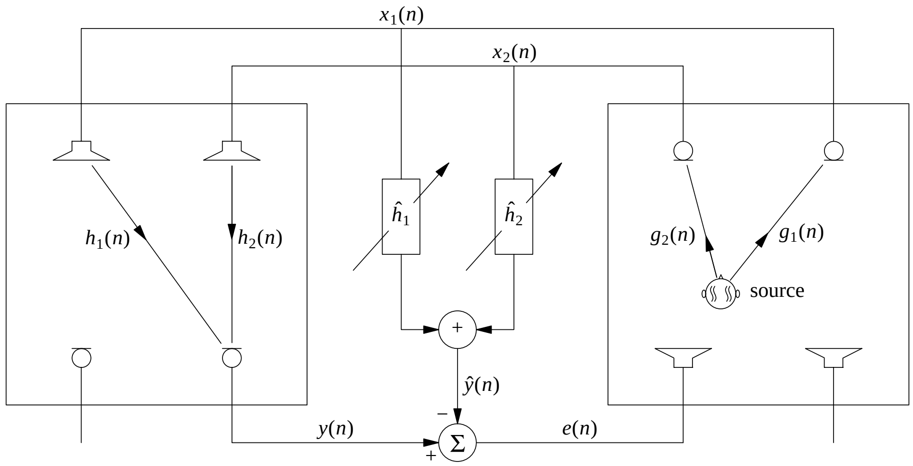

Fig. 1. Schematic diagram of stereophonic acoustic echo cancellation.

2. The extended least mean square (ELMS) algorithm

We recall here the ELMS, which is a direct approximation of the RLS algorithm; for more details see[3]. The error signal at time n is given by

$$e(n) = y(n) - \hat{y}_1(n) - \hat{y}_2(n), \tag{1}$$

where

$$\hat{y}_i(n) = \hat{\mathbf{h}}_i^T(n-1)\mathbf{x}_i(n),\ i = 1, 2, \tag{2}$$

$$\hat{\mathbf{h}}_i(n) = \left[\begin{array}{cccc} \hat{h}_{0,i}(n) & \hat{h}_{1,i}(n) & \cdots & \hat{h}_{L-1,i}(n) \end{array}\right]^T,$$

$$\mathbf{x}_i(n) = \left[\begin{array}{cccc} x_i(n) & x_i(n-1) & \cdots & x_i(n-L+1) \end{array}\right]^T.$$

L is the length of the modeling filters. The two filters are updated as follows:

$$\hat{\mathbf{h}}_1(n) = \hat{\mathbf{h}}_1(n-1) + \mu r_1^{-1}[\mathbf{x}_1(n) - \rho r_{1,2} r_{2,2}^{-1}\mathbf{x}_2(n)]e(n), \tag{3}$$

$$\hat{\mathbf{h}}_2(n) = \hat{\mathbf{h}}_2(n-1) + \mu r_2^{-1}[\mathbf{x}_2(n) - \rho r_{2,1} r_{1,1}^{-1}\mathbf{x}_1(n)]e(n), \tag{4}$$

with

$$r_{i,j}(n) = \mathbf{x}_i^T(n)\mathbf{x}_j(n),\ i, j = 1, 2, \tag{5}$$

$$r_i(n) = r_{i,i}(n)[1 - \rho^2 k^2(n)],\ i = 1, 2, \tag{6}$$

where

$$k(n) = \frac{r_{1,2}(n)}{\sqrt{r_{1,1}(n) r_{2,2}(n)}} \tag{7}$$

is the cross-correlation coefficient. We suppose in the following that $0 < \mu < 1$ and $0 \le \rho \le 1$. This algorithm is interesting because it introduces the cross-correlation coefficient between the two input signals. That can easily be exploited, as we will see later, to derive a frequency domain adaptive filter that takes into account the coherence function between the channels at different frequencies.

3. Block version of the ELMS algorithm

In this section, we derive a block version of the ELMS, which will be the foundation of the frequency domain algorithm. Let N be an integer number. We assume that L is an integer multiple of N, i.e., $L = KN$. The block error of the ELMS is:

$$\mathbf{e}(m) = \mathbf{y}(m) - \hat{\mathbf{y}}_1(m) - \hat{\mathbf{y}}_2(m), \tag{8}$$

where m is the block time index, and

$$\begin{aligned}
\mathbf{e}(m) &= \begin{bmatrix} e(mN) & \cdots & e(mN+N-2) & e(mN+N-1) \end{bmatrix}^T, \\
\mathbf{y}(m) &= \begin{bmatrix} y(mN) & \cdots & y(mN+N-2) & y(mN+N-1) \end{bmatrix}^T, \\
\hat{\mathbf{y}}_j(m) &= \begin{bmatrix} \mathbf{x}_j(mN) & \cdots & \mathbf{x}_j(mN+N-1) \end{bmatrix}^T \hat{\mathbf{h}}_j(m-1).
\end{aligned}$$

We can easily show that[4,5]

$$\hat{\mathbf{y}}_j(m) = \sum_{i=0}^{K-1} \mathbf{T}_{i,j}(m)\hat{\mathbf{h}}_{i,j}(m-1), \; j = 1, 2, \tag{9}$$

where

$$\mathbf{T}_{i,j}(m) = \begin{bmatrix} x_j(mN - Ni) & \cdots & x_j(mN - Ni - N + 1) \\ \cdots & \cdots & \cdots \\ x_j(mN - Ni + N - 1) & \cdots & x_j(mN - Ni) \end{bmatrix}$$

is an $(N \times N)$ Toeplitz matrix and

$$\hat{\mathbf{h}}_{i,j} = \begin{bmatrix} \hat{h}_{Ni,j} & \hat{h}_{Ni+1,j} & \cdots & \hat{h}_{Ni+N-1,j} \end{bmatrix}^T, \; i = 0, 1, ..., K-1, \; j = 1, 2,$$

are the sub-filters of $\hat{\mathbf{h}}_1$ and $\hat{\mathbf{h}}_2$. Suppose for now that the coefficients $r_{i,j}$ are fixed, so that the filters are adapted according to

$$\hat{\mathbf{h}}_{i,1}(m) = \hat{\mathbf{h}}_{i,1}(m-1) + \mu_{\mathrm{b}} r_1^{-1}[\mathbf{T}_{i,1}^T(m) - \rho r_{1,2} r_{2,2}^{-1} \mathbf{T}_{i,2}^T(m)]\mathbf{e}(m), \tag{10}$$

$$\hat{\mathbf{h}}_{i,2}(m) = \hat{\mathbf{h}}_{i,2}(m-1) + \mu_{\mathrm{b}} r_2^{-1}[\mathbf{T}_{i,2}^T(m) - \rho r_{2,1} r_{1,1}^{-1} \mathbf{T}_{i,1}^T(m)]\mathbf{e}(m). \tag{11}$$

From this formulation, the frequency domain adaptive algorithm is straightforward.

4. The proposed algorithm

It is well known that by doubling its size a Toeplitz matrix $\mathbf{T}$ can be transformed to a circulant matrix

$$\mathbf{C} = \begin{bmatrix} \mathbf{T}' & \mathbf{T} \\ \mathbf{T} & \mathbf{T}' \end{bmatrix}$$

where $\mathbf{T}'$ is also a Toeplitz matrix. Using circulant matrices, the block ELMS can be rewritten equivalently.
Block error signal:

$$\begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{e}(m) \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{y}(m) \end{bmatrix} - \mathbf{W}[\hat{\mathbf{y}}_1'(m) + \hat{\mathbf{y}}_2'(m)], \tag{12}$$

where

$$\mathbf{W} = \begin{bmatrix} \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \mathbf{I}_{N\times N} \end{bmatrix},$$

$$\hat{\mathbf{y}}'_j(m) = \sum_{i=0}^{K-1} \mathbf{C}_{i,j}(m) \begin{bmatrix} \hat{\mathbf{h}}_{i,j}(m-1) \\ \mathbf{0}_{N\times 1} \end{bmatrix},\ j = 1,2, \tag{13}$$

$$\mathbf{C}_{i,j}(m) = \begin{bmatrix} \mathbf{T}'_{i,j}(m) & \mathbf{T}_{i,j}(m) \\ \mathbf{T}_{i,j}(m) & \mathbf{T}'_{i,j}(m) \end{bmatrix},$$

and

$$\mathbf{T}'_{i,j}(m) = \begin{bmatrix} x_j(mN-Ni-N) & \cdots & x_j(mN-Ni+1) \\ \cdots & \cdots & \cdots \\ x_j(mN-Ni-1) & \cdots & x_j(mN-Ni-N) \end{bmatrix}.$$

Adaptation:

$$\begin{bmatrix} \hat{\mathbf{h}}_{i,1}(m) \\ \mathbf{0}_{N\times 1} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{h}}_{i,1}(m-1) \\ \mathbf{0}_{N\times 1} \end{bmatrix} + \mu_{\mathrm{b}}\mathbf{W}r_1^{-1}[\mathbf{C}_{i,1}^T(m) - \rho r_{1,2}r_{2,2}^{-1}\mathbf{C}_{i,2}^T(m)] \begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{e}(m) \end{bmatrix}, \tag{14}$$

$$\begin{bmatrix} \hat{\mathbf{h}}_{i,2}(m) \\ \mathbf{0}_{N\times 1} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{h}}_{i,2}(m-1) \\ \mathbf{0}_{N\times 1} \end{bmatrix} + \mu_{\mathrm{b}}\mathbf{W}r_2^{-1}[\mathbf{C}_{i,2}^T(m) - \rho r_{2,1}r_{1,1}^{-1}\mathbf{C}_{i,1}^T(m)] \begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{e}(m) \end{bmatrix}. \tag{15}$$

It is also well known that a circulant matrix is easily decomposed as follows: $\mathbf{C} = \mathbf{F}^{-1}\mathbf{D}\mathbf{F}$, where $\mathbf{F}$ is the Fourier matrix and $\mathbf{D}$ is a diagonal matrix whose elements are the Fourier transform of the first column of $\mathbf{C}$. Using this decomposition and the following time shift relationship[4]

$$\mathbf{D}_{i,j}(m) = \mathbf{D}_{i-1,j}(m-1),\ i = 1,2,...,K,\ j = 1,2, \tag{16}$$

(where $\mathbf{D}_{i,j}(m) = \mathbf{F}\mathbf{C}_{i,j}(m)\mathbf{F}^{-1}$) we deduce an efficient frequency domain adaptive algorithm:
filtering:

$$\underline{\hat{\mathbf{y}}}'_j(m) = \sum_{i=0}^{K-1} \mathbf{D}_{0,j}(m-i)\underline{\hat{\mathbf{h}}}_{i,j}(m-1),\ j = 1,2. \tag{17}$$

error signal:

$$\underline{\mathbf{e}}(m) = \underline{\mathbf{y}}(m) - \mathbf{F}\mathbf{W}\mathbf{F}^{-1}[\underline{\hat{\mathbf{y}}}'_1(m) + \underline{\hat{\mathbf{y}}}'_2(m)]. \tag{18}$$

adaptation:

$$\underline{\hat{\mathbf{h}}}_{i,1}(m) = \underline{\hat{\mathbf{h}}}_{i,1}(m-1) + \mu_{\mathrm{b}}\mathbf{F}\mathbf{W}\mathbf{F}^{-1}\mathbf{S}_1^{-1}[\mathbf{D}_{i,1}^*(m) - \rho\mathbf{S}_{1,2}\mathbf{S}_{2,2}^{-1}\mathbf{D}_{i,2}^*(m)]\underline{\mathbf{e}}(m), \tag{19}$$

$$\underline{\hat{\mathbf{h}}}_{i,2}(m) = \underline{\hat{\mathbf{h}}}_{i,2}(m-1) + \mu_{\mathrm{b}}\mathbf{F}\mathbf{W}\mathbf{F}^{-1}\mathbf{S}_2^{-1}[\mathbf{D}_{i,2}^*(m) - \rho\mathbf{S}_{2,1}\mathbf{S}_{1,1}^{-1}\mathbf{D}_{i,1}^*(m)]\underline{\mathbf{e}}(m), \tag{20}$$

where we have used the frequency domain quantities:

$$\begin{aligned}
\underline{\hat{\mathbf{y}}}'_j(m) &= \mathbf{F}\hat{\mathbf{y}}'_j(m), \\
\underline{\hat{\mathbf{h}}}_{i,j}(m) &= \mathbf{F}\begin{bmatrix} \hat{\mathbf{h}}_{i,j}(m) \\ \mathbf{0}_{N\times 1} \end{bmatrix}, \\
\underline{\mathbf{e}}(m) &= \mathbf{F}\begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{e}(m) \end{bmatrix}, \\
\underline{\mathbf{y}}(m) &= \mathbf{F}\begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{y}(m) \end{bmatrix}.
\end{aligned}$$

Note that if $\mathbf{S}_i = r_i\mathbf{I}_{2N\times 2N}$ and $\mathbf{S}_{i,j} = r_{i,j}\mathbf{I}_{2N\times 2N}$, then this algorithm is strictly equivalent to the block ELMS. For $\rho = 0$, we find the basic two-channel *multidelay filter* (MDF).[5]

The advantage of this approach is that the coefficients of the diagonal matrices $\mathbf{S}_i$, $i = 1, 2$, can be computed individually according to the power spectrum ($\mathbf{S}_{i,i}$), cross-power spectrum ($\mathbf{S}_{i,j}$), and coherence in the corresponding frequency bin. Hence:

$$\mathbf{S}_{i,j}(m) = \beta\mathbf{S}_{i,j}(m-1) + (1-\beta)\mathbf{D}^*_{0,i}(m)\mathbf{D}_{0,j}(m),\ i,j = 1,2, \qquad (21)$$

$$\mathbf{S}_i(m) = \mathbf{S}_{i,i}(m)[\mathbf{I}_{2N\times 2N} - \rho^2\mathbf{U}^*(m)\mathbf{U}(m)],\ i = 1,2, \qquad (22)$$

where β is a smoothing factor and

$$\mathbf{U}(m) = [\mathbf{S}_{1,1}(m)\mathbf{S}_{2,2}(m)]^{-1/2}\mathbf{S}_{1,2}(m) \qquad (23)$$

is the (diagonal) coherence matrix. It is interesting to compare Eq. (22) and Eq. (6). This algorithm has a strong link to the two-channel RLS.[6] There are much more sophisticated ways to estimate Eq. (21). One is to use the Welch method, which gives very good results in practice with speech signals.[5]

Arithmetic complexity:
Suppose that the block length N is a power of 2, i.e., $N = 2^b$. We also assume that the FFT is computed using the split radix algorithm.[7] Then the number of operations to be performed per output point for the proposed algorithm is roughly: $4K(b+12)+4b$ real multiplications and $4K(3b+10)+12b$ real additions. For example with $L = 1024$ and $N = 256$ the computational cost of the proposed algorithm is about 8 times smaller than the two-channel NLMS.

The proposed algorithm can be easily generalized to the *generalized multidelay filter* (GMDFα),[8] where α is the overlap factor. This structure is very useful in the context of adaptive filtering, because the filter coefficients are adapted more frequently (every $M = N/\alpha$ samples instead of every N samples). As a result, a faster convergence rate and a better tracking are expected. However, the complexity is increased by a factor α. Of course, we can also derive an unconstrained version to reduce the complexity.[9]

5. Conclusions

We presented a new and efficient (at least in theory) frequency domain adaptive algorithm exploiting the coherence between the input channels to reduce the detrimental effect of their correlation. The low complexity and the good behavior of the proposed algorithm will probably make it a very good candidate for stereophonic and, more generally, multichannel acoustic echo cancellation.

References

[1] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic acoustic echo cancellation—An overview of the fundamental problem," IEEE Signal Processing Lett. **2**, 148-151 (1995).

[2] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," IEEE Trans. Speech Audio Processing **6**, 156-165 (1998).

[3] J. Benesty, F. Amand, A. Gilloire, and Y. Grenier, "Adaptive filtering algorithms for stereophonic acoustic echo cancellation," in *Proc. IEEE ICASSP, 1995*, pp. 3099-3102.

[4] J. Benesty and P. Duhamel, "Fast constant modulus adaptive algorithm," IEE Proc.-F **138**, 379-387 (1991).

[5] J.-S. Soo and K. K. Pang, "Multidelay block frequency domain adaptive filter," IEEE Trans. Acoust., Speech, Signal Processing **38**, 373-376 (1990).

[6] J. Benesty, P. Duhamel, and Y. Grenier, "Multi-channel adaptive filtering applied to multi-channel acoustic echo cancellation," in *Proc. EUSIPCO, 1996*.

[7] P. Duhamel, "Implementation of split-radix FFT algorithm for complex, real and real-symmetric data," IEEE Trans. Acoust., Speech, Signal Processing **34**, 285-295 (1986).

[8] E. Moulines, O. Ait Amrane, and Y. Grenier, "The generalized multidelay adaptive filter: structure and convergence analysis," IEEE Trans. Signal Processing **43**, 14-28 (1995).

[9] D. Mansour and A. H. Gray, JR., "Unconstrained frequency-domain adaptive filter," IEEE Trans. Acoust., Speech, Signal Processing **30**, 726-734 (1982).

