EURASIP Journal on Applied Signal Processing 2003:11, 1056–1063

A Multidelay Double-Talk Detector Combined with the MDF Adaptive Filter

Jacob Benesty
Université du Québec, INRS-EMT 800 de la Gauchetière Ouest, Suite 6900 Montréal, Québec, Canada H5A 1K6
Email: benesty@inrs-emt.uquebec.ca

Tomas Gänsler
Agere Systems, 555 Union Boulevard, Allentown, PA 18109-3229, USA
Email: gaensler@agere.com

Received 31 July 2002 and in revised form 5 March 2003

The multidelay block frequency-domain (MDF) adaptive filter is an excellent candidate for both acoustic and network echo cancellation. There is a need for a very good double-talk detector (DTD) to be combined efficiently with the MDF algorithm. Recently, a DTD based on a normalized cross-correlation vector was proposed and it was shown that this DTD performs much better than the Geigel algorithm and other DTDs based on the cross-correlation coefficient. In this paper, we show how to extend the definition of a normalized cross-correlation vector in the frequency domain for the general case where the block size of the Fourier transform is smaller than the length of the adaptive filter. The resulting DTD has an MDF structure, which makes it easy to implement, and a good fit with an echo canceler based on the MDF algorithm. We also analyze resource requirements (computational complexity and memory requirement) and compare the MDF algorithm with the normalized least mean square algorithm (NLMS) from this point of view.

Keywords and phrases: adaptive filtering, frequency domain, double-talk detection, echo cancellation.

1. INTRODUCTION

Network and acoustic echo cancelers work on the same principle. An echo canceler (EC) [1], to work well, should include good solutions to two important problems: a system identification problem and a so-called *double-talk* detection problem [2]. When the echo path is identified by an adaptive filter, a function should be included to freeze the adaptation whenever a near-end signal is detected, and thereby avoid the divergence of the adaptive algorithm. This control can either be done by a so-called step-size control (soft decision) or by a double-talk detector (DTD) hard decision. Theoretically, the step-size control method would be preferable because it can be made optimal in minimum mean-square sense [3, 4, 5]. In practice however, depending on situation, there is no conclusive evidence that soft decisions (step-size control) result in better performance than using the DTD hard decisions. Hence, it is of great interest to find a suitable and practical decision variable.

One of the most widely used DTDs is the Geigel algorithm [6] which works fairly well when the echo return loss is well defined. However, this is not, in general, the case in practice. The need for more sophisticated DTDs that do not depend on the path attenuation is obvious. Alternative methods for double-talk detection have been presented, for example, in [7, 8]. A family of DTDs exhibiting this feature was proposed in [9].

On the system identification part, the multidelay block frequency-domain (MDF) adaptive filter [10] is an excellent candidate for both acoustic and network echo cancellation. Indeed, since the coefficients of this adaptive filter are updated in the frequency domain, block by block, using the fast Fourier transform (FFT) as an intermediary step, it is very efficient from a complexity point of view. Moreover, the block length N is independent of the filter length L; N can be chosen as small as desired, with a resulting algorithmic delay equal to N. Although, from a complexity point of view, the optimal choice is $N = L$, using smaller block sizes ($N < L$) in order to reduce the delay is still more efficient than time-domain algorithms. The block delay is not a problem for some applications, for example, in a frame-based system like a Voice-over-Internet Protocol (VoIP) network. In this network, even a sample-by-sample time-domain algorithm would introduce a delay equal to the delay of a block-based algorithm. Hence, there is no delay penalty using a block-based MDF algorithm in this scenario if its block size is matched to the frame size of the network.

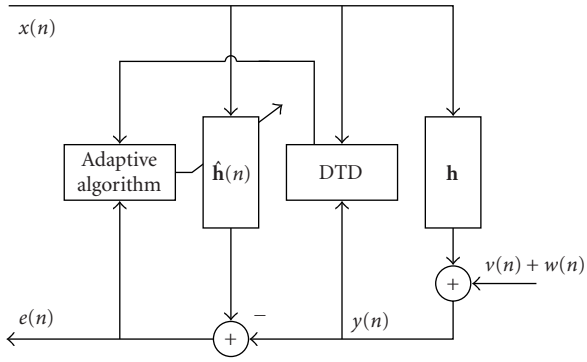


Figure 1: Block diagram of the echo canceler (EC), double-talk detector (DTD), and echo path.

A DTD based on a normalized cross-correlation vector was proposed in [9]. In [2], it was shown that this DTD performs much better than the Geigel algorithm and other DTDs based on the cross-correlation coefficient. In this paper, we show how to extend the ideas of [9] to the MDF algorithm. The resulting DTD has an MDF structure which makes it easy to implement and a good fit with an EC based on the MDF algorithm.

The organization of this paper is as follows. In Section 2, we introduce some definitions and notation that are used in the context of echo cancellation. In Section 3, we give the MDF algorithm. Section 4 presents the new DTD and its combination with an MDF EC. A resource analysis of the MDF algorithm is given in Section 5. Evaluation of the proposed MDF DTD is made in Section 6. Finally, we give our conclusions in Section 7.

2. DEFINITIONS AND NOTATION

Referring to Figure 1, the following definitions and notation are used in all the derivations:

(i) $x(n)$ = far-end signal/speech,
(ii) $w(n)$ = ambient (background) noise,
(iii) $v(n)$ = near-end signal/speech (double-talk),
(iv) $\mathbf{x}(n) = [x(n) \ \cdots \ x(n-L+1)]^T$, excitation vector,
(v) $y(n) = \mathbf{h}^T\mathbf{x}(n) + w(n) + v(n)$, that is, echo + ambient noise + near-end signal,
(vi) $\mathbf{h} = [h_0 \ \cdots \ h_{L-1}]^T$, true echo path vector,
(vii) $\hat{\mathbf{h}}(n) = [\hat{h}_0(n) \ \cdots \ \hat{h}_{L-1}(n)]^T$, estimated echo path vector,
(viii) $\hat{y}(n) = \hat{\mathbf{h}}^T(n-1)\mathbf{x}(n)$, estimated echo,
(ix) $e(n) = y(n) - \hat{y}(n)$, error signal.

Here, n is the sample-by-sample time index and L is the length of the adaptive filter that we suppose to be equal to the length of the echo path.

3. THE MDF ADAPTIVE FILTER

In this section, we give the MDF algorithm [10]. For further details and explanation, see [10, 11]. We assume that L is an integer multiple of N, that is, $L = KN$. We define the block error signal (of length $N \le L$) as

$$\mathbf{e}(m) = \mathbf{y}(m) - \hat{\mathbf{y}}(m), \tag{1}$$

where m is the block time index, and

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

The vector $\hat{\mathbf{h}}$ is defined in the same manner as $\hat{\mathbf{h}}(n)$ in the previous section. It can easily be checked that $\mathbf{X}$ is a Toeplitz matrix of size $L \times N$.

We can show that

$$\hat{\mathbf{y}}(m) = \sum_{k=0}^{K-1} \mathbf{T}(m-k)\hat{\mathbf{h}}_k, \tag{3}$$

where

$$\mathbf{T}(m-k) = \begin{bmatrix} x(mN-kN) & \cdots & x(mN-kN-N+1) \\ x(mN-kN+1) & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ x(mN-kN+N-1) & \cdots & x(mN-kN) \end{bmatrix} \tag{4}$$

is an $N \times N$ Toeplitz matrix and

$$\hat{\mathbf{h}}_k = \begin{bmatrix} \hat{h}_{kN} & \hat{h}_{kN+1} & \cdots & \hat{h}_{kN+N-1} \end{bmatrix}^T, \quad k = 0, 1, \ldots, K-1, \tag{5}$$

are the subfilters of $\hat{\mathbf{h}}$. In (3), the filter $\hat{\mathbf{h}}$ (of length L) is partitioned in K subfilters $\hat{\mathbf{h}}_k$ of length N and the rectangular matrix $\mathbf{X}^T$ (of size $N \times L$) is decomposed in K square submatrices of size $N \times N$.

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

$$\mathbf{C} = \begin{bmatrix} \mathbf{T}' & \mathbf{T} \\ \mathbf{T} & \mathbf{T}' \end{bmatrix}, \tag{6}$$

where $\mathbf{T}'$ is also a Toeplitz matrix. (The matrix $\mathbf{T}'$ is expressible in terms of the elements of $\mathbf{T}$, except for an arbitrary diagonal.) 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 (of size $2N \times 2N$) and $\mathbf{D}$ is a diagonal matrix whose elements are the discrete Fourier transform of the first column of $\mathbf{C}$.

Now, we define the frequency-domain quantities

$$\begin{aligned}\underline{\mathbf{y}}(m) &= \mathbf{F}\begin{bmatrix}\mathbf{0}_{N\times 1}\\ \mathbf{y}(m)\end{bmatrix},\\ \hat{\underline{\mathbf{h}}}_k(m) &= \mathbf{F}\begin{bmatrix}\hat{\mathbf{h}}_k(m)\\ \mathbf{0}_{N\times 1}\end{bmatrix},\\ \underline{\mathbf{e}}(m) &= \mathbf{F}\begin{bmatrix}\mathbf{0}_{N\times 1}\\ \mathbf{e}(m)\end{bmatrix}.\end{aligned} \tag{7}$$

The MDF adaptive filter is then given by the following equations:

$$\begin{aligned}\underline{\mathbf{e}}(m) &= \underline{\mathbf{y}}(m) - \mathbf{G}^{01}\sum_{k=0}^{K-1}\mathbf{D}(m-k)\hat{\underline{\mathbf{h}}}_k(m-1),\\ \mathbf{S}_{\mathrm{MDF}}(m) &= \lambda\mathbf{S}_{\mathrm{MDF}}(m-1) + (1-\lambda)\mathbf{D}^*(m)\mathbf{D}(m),\\ \hat{\underline{\mathbf{h}}}_k(m) &= \hat{\underline{\mathbf{h}}}_k(m-1) + \mu(1-\lambda)\mathbf{G}^{10}\mathbf{D}^*(m-k)\\ &\quad\times\left[\mathbf{S}_{\mathrm{MDF}}(m) + \delta\mathbf{I}_{2N\times 2N}\right]^{-1}\underline{\mathbf{e}}(m),\end{aligned} \tag{8}$$

where $k = 0, 1, \ldots, K-1$, $*$ denotes complex conjugate, λ ($0 \ll \lambda < 1$) is an exponential forgetting factor, μ ($0 < \mu \leq 2$) is a positive number, δ is a regularization parameter, and

$$\begin{aligned}\mathbf{G}^{01} &= \mathbf{F}\mathbf{W}^{01}\mathbf{F}^{-1},\\ \mathbf{W}^{01} &= \begin{bmatrix}\mathbf{0}_{N\times N} & \mathbf{0}_{N\times N}\\ \mathbf{0}_{N\times N} & \mathbf{I}_{N\times N}\end{bmatrix},\\ \mathbf{G}^{10} &= \mathbf{F}\mathbf{W}^{10}\mathbf{F}^{-1},\\ \mathbf{W}^{10} &= \begin{bmatrix}\mathbf{I}_{N\times N} & \mathbf{0}_{N\times N}\\ \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N}\end{bmatrix}.\end{aligned} \tag{9}$$

We now turn the focus of this paper on a DTD that fits well with the MDF adaptive filter. In the next section, we derive this DTD and show how to combine it with the MDF algorithm.

4. A MULTIDELAY DOUBLE-TALK DETECTOR

The best way we know to detect the presence of double talk is to form a test statistic ξ and compare it to a threshold T: if $\xi \geq T$, then we say that double talk is not present; if $\xi < T$, then we say that double talk is present. The test statistic is, in general, related to correlation or coherence and the threshold must be a known constant for best performance.

In the derivation of the DTD, we will neglect the effect of noise (e.g., $w = 0$) for simplicity. It can easily be checked that

$$\begin{aligned}\underline{\mathbf{y}}(m) &= \mathbf{G}^{01}\sum_{k=0}^{K-1}\mathbf{D}(m-k)\underline{\mathbf{h}}_k + \underline{\mathbf{v}}(m)\\ &= \mathbf{G}^{01}\underline{\mathbf{D}}(m)\underline{\mathbf{h}}_{2L} + \underline{\mathbf{v}}(m),\end{aligned} \tag{10}$$

where

$$\begin{aligned}\underline{\mathbf{D}}(m) &= \begin{bmatrix}\mathbf{D}(m) & \mathbf{D}(m-1) & \cdots & \mathbf{D}(m-K+1)\end{bmatrix},\\ \underline{\mathbf{h}}_{2L} &= \begin{bmatrix}\underline{\mathbf{h}}_0^T & \underline{\mathbf{h}}_1^T & \cdots & \underline{\mathbf{h}}_{K-1}^T\end{bmatrix}^T,\\ \underline{\mathbf{h}}_k &= \mathbf{F}\begin{bmatrix}\mathbf{h}_k\\ \mathbf{0}_{N\times 1}\end{bmatrix},\\ \mathbf{v}(m) &= \begin{bmatrix}v(mN) & \cdots & v(mN+N-1)\end{bmatrix}^T,\\ \underline{\mathbf{v}}(m) &= \mathbf{F}\begin{bmatrix}\mathbf{0}_{N\times 1}\\ \mathbf{v}(m)\end{bmatrix}.\end{aligned} \tag{11}$$

Suppose that $v = 0$. In this case,

$$\sigma_y^2 = E\{\underline{\mathbf{y}}^H(m)\underline{\mathbf{y}}(m)\} = \underline{\mathbf{h}}_{2L}^H\mathbf{S}\underline{\mathbf{h}}_{2L}, \tag{12}$$

where H denotes conjugate transpose, $E\{\cdot\}$ is the mathematical expectation, and

$$\mathbf{S} = E\{\underline{\mathbf{D}}^H(m)\mathbf{G}^{01}\underline{\mathbf{D}}(m)\}. \tag{13}$$

Thanks to (10) and (13), we have

$$E\{\underline{\mathbf{D}}^H(m)\underline{\mathbf{y}}(m)\} = \mathbf{S}\underline{\mathbf{h}}_{2L} = \mathbf{s}, \tag{14}$$

and (12) can be rewritten as

$$\sigma_y^2 = \underline{\mathbf{h}}_{2L}^H\mathbf{s} = \sum_{k=0}^{K-1}\underline{\mathbf{h}}_k^H E\{\mathbf{D}^*(m-k)\underline{\mathbf{y}}(m)\} = \sum_{k=0}^{K-1}\underline{\mathbf{h}}_k^H\mathbf{s}_k, \tag{15}$$

with

$$\mathbf{s}_k = E\{\mathbf{D}^*(m-k)\underline{\mathbf{y}}(m)\}. \tag{16}$$

Now, in general, for $v \neq 0$,

$$\sigma_y^2 = \underline{\mathbf{h}}_{2L}^H\mathbf{s} + \sigma_v^2, \tag{17}$$

where

$$\sigma_v^2 = E\{\underline{\mathbf{v}}^H(m)\underline{\mathbf{v}}(m)\}. \tag{18}$$

Basically, there are two different ways to compute σ_y^2 when no double talk is present, and we take advantage of this information to detect the presence of a near-end signal. If we divide (15) by (17), we obtain the following decision variable:

$$\xi^2 = \frac{\underline{\mathbf{h}}_{2L}^H\mathbf{s}}{\underline{\mathbf{h}}_{2L}^H\mathbf{s} + \sigma_v^2} = \frac{\eta_y^2}{\sigma_y^2}. \tag{19}$$

We easily deduce from (19) that for $v = 0$, $\xi = 1$, and for $v \neq 0$, $\xi < 1$. Note also that ξ is not, in principle, sensitive to changes of the echo path when $v = 0$.

In practice, ξ is estimated recursively as follows:

$$\xi^2(m) = \frac{\sum_{k=0}^{K-1}\hat{\underline{\mathbf{h}}}_{\mathrm{b},k}^H(m)\mathbf{s}_k(m)}{\sigma_y^2(m)} = \frac{\eta_y^2(m)}{\sigma_y^2(m)}. \tag{20}$$

- Spectral and correlation estimation

$$\mathbf{S}_{\mathrm{MDF}}(m) = \lambda \mathbf{S}_{\mathrm{MDF}}(m-1) + (1-\lambda)\mathbf{D}^*(m)\mathbf{D}(m)$$

$$\sigma_y^2(m) = \lambda_{\mathrm{b}} \sigma_y^2(m-1) + (1-\lambda_{\mathrm{b}})\underline{\mathbf{y}}^H(m)\underline{\mathbf{y}}(m)$$

$$\mathbf{s}_k(m) = \lambda_{\mathrm{b}} \mathbf{s}_k(m-1) + (1-\lambda_{\mathrm{b}})\mathbf{D}^*(m-k)\underline{\mathbf{y}}(m)$$

- MDF DTD (background filter)

$$\underline{\mathbf{e}}_{\mathrm{b}}(m) = \underline{\mathbf{y}}(m) - \mathbf{G}^{01} \sum_{k=0}^{K-1} \mathbf{D}(m-k)\hat{\underline{\mathbf{h}}}_{\mathrm{b},k}(m-1)$$

$$\hat{\underline{\mathbf{h}}}_{\mathrm{b},k}(m) = \hat{\underline{\mathbf{h}}}_{\mathrm{b},k}(m-1) + (1-\lambda_{\mathrm{b}})\mathbf{G}^{10}\mathbf{D}^*(m-k)\left[\mathbf{S}_{\mathrm{MDF}}(m) + \delta \mathbf{I}_{2N\times 2N}\right]^{-1}\underline{\mathbf{e}}_{\mathrm{b}}(m)$$

$$\xi^2(m) = \frac{\sum_{k=0}^{K-1} \hat{\underline{\mathbf{h}}}_{\mathrm{b},k}^H(m)\mathbf{s}_k(m)}{\sigma_y^2(m)}$$

$$\xi(m) < T \Longrightarrow \text{double talk, } \mu = 0$$

$$\xi(m) \geq T \Longrightarrow \text{no double talk, } \mu$$

- MDF EC (foreground filter)

$$\underline{\mathbf{e}}(m) = \underline{\mathbf{y}}(m) - \mathbf{G}^{01} \sum_{k=0}^{K-1} \mathbf{D}(m-k)\hat{\underline{\mathbf{h}}}_k(m-1)$$

$$\hat{\underline{\mathbf{h}}}_k(m) = \hat{\underline{\mathbf{h}}}_k(m-1) + \mu(1-\lambda)\mathbf{G}^{10}\mathbf{D}^*(m-k)\left[\mathbf{S}_{\mathrm{MDF}}(m) + \delta \mathbf{I}_{2N\times 2N}\right]^{-1}\underline{\mathbf{e}}(m)$$

Scheme 1: The MDF adaptive filter combined with a multidelay DTD.

The echo path of the system is estimated, in the test statistic, by a background MDF adaptive filter $\hat{\underline{\mathbf{h}}}_{\mathrm{b},k}$, $k = 0, 1, \ldots, K-1$, with an exponential window λ_{b} ($0 \ll \lambda_{\mathrm{b}} < 1$) smaller than λ, the exponential window used for the system identification by a foreground MDF algorithm. However, what is important in practice is that the statistics of the signal $y(n)$ (containing both the echo and the near-end signal during double talk) is tracked fast enough, faster than the statistics of the update of the foreground filter, hence λ_{b} is chosen smaller than λ. We have to use $\mu = 1$ for the background filter so that the two different ways we compute the statistics of $y(n)$ (numerator and denominator of (19)) are consistent and estimated at the same rate. This way, the DTD alerts the foreground filter before it diverges by freezing its adaptation during double-talk. Furthermore, for practical reasons, even though not mathematically stringent, we use the same spectral matrix $\mathbf{S}_{\mathrm{MDF}}(m)$ for the foreground and background filters. All the variables used in the test statistic are estimated as

$$\begin{aligned}
\mathbf{s}_k(m) &= \lambda_{\mathrm{b}}\mathbf{s}_k(m-1) + (1-\lambda_{\mathrm{b}})\mathbf{D}^*(m-k)\underline{\mathbf{y}}(m),\\
\sigma_y^2(m) &= \lambda_{\mathrm{b}}\sigma_y^2(m-1) + (1-\lambda_{\mathrm{b}})\underline{\mathbf{y}}^H(m)\underline{\mathbf{y}}(m),\\
\underline{\mathbf{e}}_{\mathrm{b}}(m) &= \underline{\mathbf{y}}(m) - \mathbf{G}^{01}\sum_{k=0}^{K-1}\mathbf{D}(m-k)\hat{\underline{\mathbf{h}}}_{\mathrm{b},k}(m-1),\\
\hat{\underline{\mathbf{h}}}_{\mathrm{b},k}(m) &= \hat{\underline{\mathbf{h}}}_{\mathrm{b},k}(m-1) + (1-\lambda_{\mathrm{b}})\mathbf{G}^{10}\mathbf{D}^*(m-k)\\
&\quad\times\left[\mathbf{S}_{\mathrm{MDF}}(m)+\delta\mathbf{I}_{2N\times 2N}\right]^{-1}\underline{\mathbf{e}}_{\mathrm{b}}(m),
\end{aligned} \tag{21}$$

where $k = 0, 1, \ldots, K-1$.

Scheme 1 summarizes the combination of the MDF EC and the MDF DTD, where $k = 0, 1, \ldots, K-1$; $0 < \mu \leq 2$ is an adaptation step; λ, λ_{b} are exponential windows; δ is the regularization factor; T is the threshold,

$$\begin{aligned}
\mathbf{G}^{01} &= \mathbf{F}\mathbf{W}^{01}\mathbf{F}^{-1}, \qquad \mathbf{W}^{01} = \begin{bmatrix} \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \mathbf{I}_{N\times N} \end{bmatrix},\\
\mathbf{G}^{10} &= \mathbf{F}\mathbf{W}^{10}\mathbf{F}^{-1}, \qquad \mathbf{W}^{10} = \begin{bmatrix} \mathbf{I}_{N\times N} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} \end{bmatrix}.
\end{aligned} \tag{22}$$

Next, we will take a look at the numerical complexity and memory requirement of the core MDF algorithm.

5. RESOURCE ANALYSIS OF THE MDF ADAPTIVE FILTER

An arithmetic operation (op.) is considered to be any real multiplication, real addition, real subtraction, or real division. Assume that

$$z_1 = a + jb, \qquad z_2 = c + jd. \tag{23}$$

Complex operations are transformed into real operations according to Table 1.

A complex variable is assumed to require two memory locations. For a Fourier-transformed vector, we assume that

Table 1

Complex operations	Real multiplications	Real additions
$z_1 \cdot z_2 = (a + jb)(c + jd)$ $= ac - bd + j(ad + bc)$	4	2
$z_1 \pm z_2 = (a + jb) \pm (c + jd)$ $= (a \pm c) + j(b \pm d)$	0	2

only half its elements need to be stored, that is, the memory required for a vector of length N is equivalent in both time and frequency domains. If a Fourier transform of length N is computed using the FFT routine devised by [12], it requires

$$\text{Mult} : \frac{N}{2}\log_2[N] - \frac{5N}{4},$$
$$\text{Add} : \frac{3N}{2}\log_2[N] - \frac{N}{4} - 4,$$
$$\text{Total op.} : 2N\log_2[N] - \frac{3N}{2} - 4.$$

As a reference, we will use the real-valued NLMS algorithm [13] (assuming all signals are real-valued) which is the workhorse algorithm of network ECs. Tables 2 and 3 show the resource requirements for the MDF and the basic real-valued NLMS algorithms with respect to their computational complexity and memory. In Figure 2, these requirements are compared, with a filter length of $L = 512$ and various block sizes N. The trade-off between computational and memory requirements is clearly exemplified. These values, however, do not translate directly to complexity for a specific hardware, but are meant to give a more general insight to required resources.

6. SIMULATIONS

In this section, we present some performance results in the context of network echo cancellation. Figure 1 shows the principle of a network EC. The far-end speech signal $x(n)$ goes through the echo path represented by a filter $\mathbf{h}$, then it is added to the near-end talker signal $v(n)$ and the ambient noise $w(n)$. The composite signal is denoted by $y(n)$. Most often, the echo path is modeled by an adaptive FIR filter $\hat{\mathbf{h}}(n)$ which subtracts a replica of the echo and thereby achieves cancellation. Double talk occurs when the two talkers on both sides speak simultaneously, that is, $x(n) \neq 0$ and $v(n) \neq 0$. In this situation, the near-end speech acts as a high-level uncorrelated noise to the adaptive algorithm. The disturbing near-end speech may therefore cause the adaptive filter to diverge, passing annoying audible echo to the far end. A common way to alleviate this problem is to slow down or completely halt the filter adaptation when near-end speech is detected. This is the very important role of the DTD. Figure 3 shows a typical network impulse response that we have used

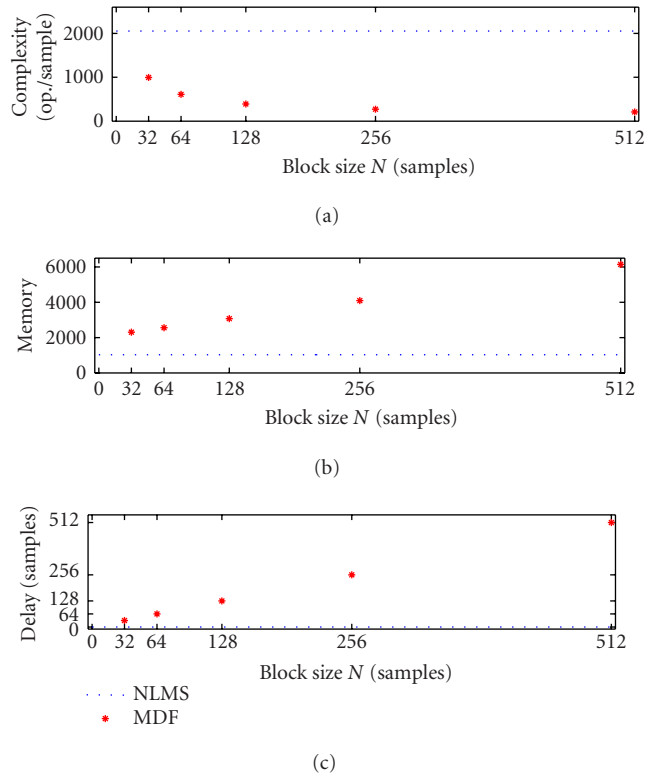


Figure 2: Resource requirement comparison of full-band (real-valued) NLMS and MDF adaptive filter designs for $L = 512$, see Table 2 for general L and N. (a) Required operations/sample. (b) Required memory locations. (c) Algorithmic delay.

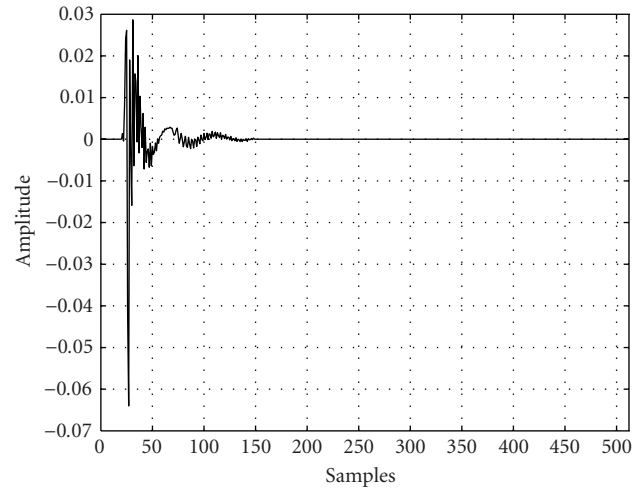


Figure 3: Impulse response used in simulations.

in all our simulations. Even though the active coefficients in this case occur in the early part of the impulse response, it is not the case in general. Hence, in this application, we always have to cover a longer time span than the active region. The time span of this network echo path $\mathbf{h}$ is 64 milliseconds ($L = 512$). The same length is used for the adaptive filter

Table 2: Complexity and memory requirements for the MDF algorithm. The computations in this version are slightly reorganized, compared to the ones in Scheme 1.

Algorithm step	Operations	Memory
$\mathbf{D}(m) = \operatorname{diag}\left\{\mathbf{F}\begin{bmatrix} x(mN-N) \\ \vdots \\ x(mN+N-1)\end{bmatrix}\right\}$	$4N\log_2[2N] - 3N - 4$	$2L + 2N$
$\mathbf{y}(m) = \begin{bmatrix} y(mN-N+1) & \cdots & y(mN) & \mathbf{0}_{1\times N}\end{bmatrix}^T$	0	N
$\mathbf{S}_{\mathrm{MDF}}(m) = \lambda\mathbf{S}_{\mathrm{MDF}}(m-1) + \mathbf{D}^*(m)\mathbf{D}(m)$	$5N$	N
$\mathbf{e}(m) = \mathbf{y}(m) - \mathbf{W}^{01}\mathbf{F}^{-1}\sum_{k=0}^{K-1}\mathbf{D}(m-k)\hat{\underline{\mathbf{h}}}_k(m-1)$	$6L - 2N + 4N\log_2[2N] - 4$	N
$\underline{\mathbf{e}}(m) = \mathbf{F}\begin{bmatrix}\mathbf{e}^T(m) & \mathbf{0}_{1\times N}\end{bmatrix}^T$	$4N\log_2[2N] - 3N - 4$	$2N$
$\mathbf{S}_{\mathrm{reg.}}(m) = \mathbf{S}_{\mathrm{MDF}}(m) + \delta\mathbf{I}_{2N\times 2N}$	N	N
$\hat{\underline{\mathbf{h}}}_k(m) = \hat{\underline{\mathbf{h}}}_k(m-1) + \mu\mathbf{G}^{10}\mathbf{S}^{-1}_{\mathrm{reg.}}(m)\mathbf{D}^*(m-k)\underline{\mathbf{e}}(m)$ $k = 0, 1, \ldots, K-1$	$4L + 2N + 8L\log_2[2N] - 8K$	$2L$
Total	$10L - 8K - 12 + 4(2L+3N)\log_2[2N]$	$4L + 8N$
Total/sample	$\frac{10L}{N} - \frac{8K}{N} - \frac{12}{N} + \frac{4(2L+3N)}{N}\log_2[2N]$	$4L + 8N$

Table 3: Complexity and memory requirements for the (real-valued) NLMS algorithm.

Algorithm step	Operations	Memory
$P_x(-1) = \delta$	0	1
$\mathbf{x}(n) = \begin{bmatrix} x(n) & \cdots & x(n-L+1)\end{bmatrix}^T$		L
$P_x(n) = P_x(n-1) + x^2(n) - x^2(n-L)$	4	1
$e(n) = y(n) - \hat{\mathbf{h}}^T\mathbf{x}(n)$	$2L$	1
$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \frac{\mu}{P_x(n)}\mathbf{x}(n)e(n)$	$2L + 3$	L
Total/sample	$4L + 7$	$2L + 3$

$\hat{\mathbf{h}}(n)$. The far-end speaker is a female (Figure 4a) and the near-end speaker is a male (Figure 4b). The sampling rate is 8 kHz and the echo-to-ambient-noise ratio is equal to 39 dB. The following parameters are used for the algorithms:

$$\begin{aligned} &N = 128, \\ &\mu = 2, \quad \lambda = \left(1 - \frac{1}{3L}\right)^N, \\ &T = 0.91, \quad \lambda_{\mathrm{b}} = \left(1 - \frac{2}{3L}\right)^N, \\ &\hat{\underline{\mathbf{h}}}_{\mathrm{b},k}(0) = \hat{\underline{\mathbf{h}}}_k(0) = \mathbf{0}. \end{aligned} \tag{24}$$

Performance is measured by means of the normalized misalignment defined as

$$\frac{||\mathbf{h} - \hat{\mathbf{h}}(n)||^2}{||\mathbf{h}||^2}. \tag{25}$$

Figure 4c shows the misalignment of the MDF EC when combined with the proposed DTD. Double talk starts around 1.3 seconds. We can see that the proposed MDF DTD detects quickly the near-end signal and freezes the adaptation of the (foreground) adaptive filter during the whole time of double talking. Of course without a DTD, the algorithm would have diverged very quickly.

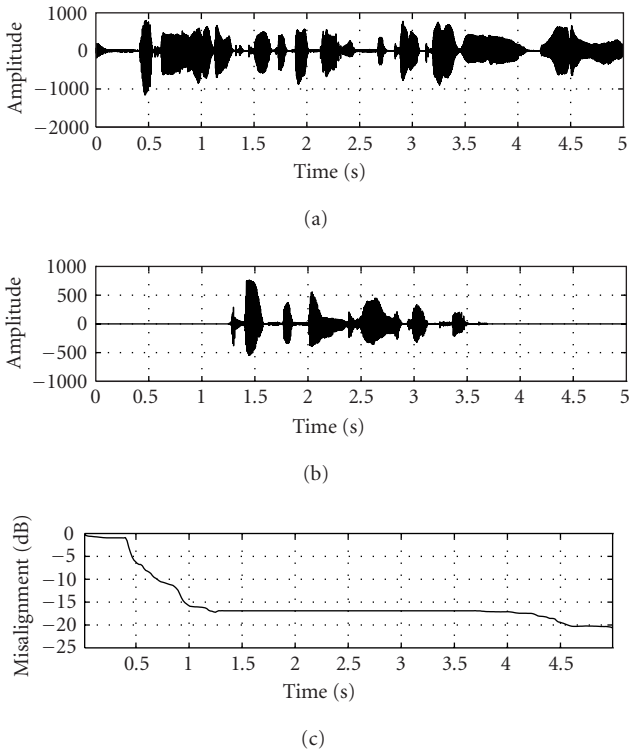


Figure 4: Behavior, during double-talk situation, of the MDF EC when combined with the proposed MDF DTD. (a) Far-end signal. (b) Near-end signal. (c) Misalignment of the MDF EC.

Figure 5 shows the performance of the EC after an abrupt system change where the impulse response is shifted 200 samples in 2 seconds. In this simulation, there is no double talk. Figure 5a (respectively, Figure 5b) corresponds to the case where the MDF DTD is deactivated (respectively, activated). We can see that the performance of the EC with the MDF DTD is slightly degraded than without. This is due to the fact that any DTD will trigger false alarms; consequently, adaptation is frozen during that time and convergence slows down. This unideal behavior is mainly caused by short-term correlation of the statistics used in the DTD. However, it has been shown that the false alarm rate of the proposed DTD structure is in general considerably lower than that of the Geigel DTD [14].

7. CONCLUSIONS

Double-talk detection is an important part of an EC system. A good DTD should be able to distinguish between double talk and echo path changes, and the threshold T should be a known constant. In this paper, we have proposed a new DTD that has these features by extending the definition of a normalized cross-correlation vector [9] in the frequency domain for the general case $N \leq L$. Purposely, the proposed DTD has an MDF structure in order to take advantage of the good characteristics of the MDF algorithm and to make a successful integration between the MDF DTD and an MDF EC.

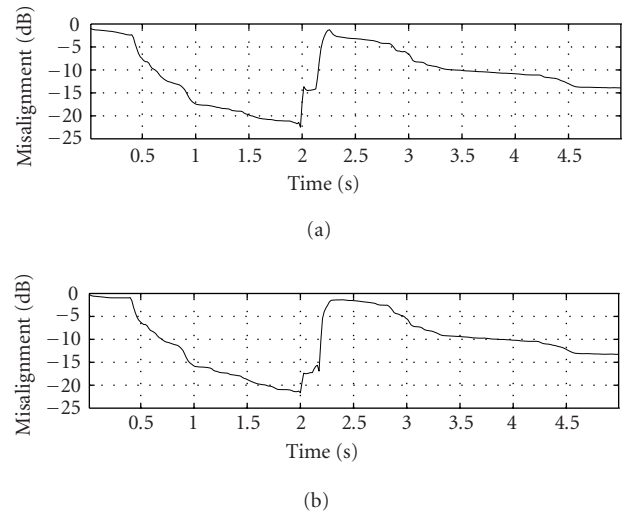


Figure 5: Convergence and tracking of the MDF EC when the MDF DTD is (a) deactivated and (b) activated.

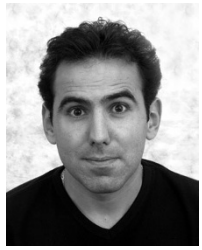
With the MDF algorithm, we can easily trade off computational load with memory requirement and algorithmic delay, hence tailor the algorithm for a specific application. For example, in a frame-based VoIP system, no delay penalty is introduced compared to a time-domain (zero-delay) algorithm as long as the block size is matched to the frame size.

We can also use robust statistics [15] to derive a robust MDF adaptive filter, the same way it was done in [11] for the FLMS algorithm ($N = L$). A robust algorithm permits decreasing the threshold T without losing performance during double-talk; as a result, the probability of false alarm is low and the performance (convergence and tracking) of the adaptive algorithm is not much affected.

REFERENCES

[1] M. M. Sondhi, "An adaptive echo canceler," *Bell System Technical Journal*, vol. 46, no. 3, pp. 497–511, 1967.
[2] J. H. Cho, D. R. Morgan, and J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," *IEEE Trans. Speech, and Audio Processing*, vol. 7, no. 6, pp. 718–724, 1999.
[3] S. Yamamoto and S. Kitayama, "An adaptive echo canceller with variable step gain method," *Trans. IECE Japan*, vol. E 65, no. 1, pp. 1–8, 1982.
[4] C. Breining, P. Dreiseitel, E. Hänsler, et al., "Acoustic echo control. An application of very-high-order adaptive filters," *IEEE Signal Processing Magazine*, vol. 16, no. 4, pp. 42–69, 1999.
[5] A. Mader, H. Puder, and G. U. Schmidt, "Step-size control for acoustic echo cancellation filters—an overview," *Signal Processing*, vol. 80, no. 9, pp. 1697–1719, 2000.
[6] D. L. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Trans. Communications*, vol. 26, pp. 647–653, May 1978.
[7] H. Ye and B.-X. Wu, "A new double-talk detection algorithm based on the orthogonality theorem," *IEEE Trans. Communications*, vol. 39, no. 11, pp. 1542–1545, 1991.
[8] T. Gänsler, M. Hansson, C.-J. Ivarsson, and G. Salomonsson, "A double-talk detector based on coherence," *IEEE Trans. Communications*, vol. 44, no. 11, pp. 1421–1427, 1996.

[9] J. Benesty, D. R. Morgan, and J. H. Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Trans. Speech, and Audio Processing*, vol. 8, no. 2, pp. 168–172, 2000.
[10] J.-S. Soo and K. K. Pang, "Multidelay block frequency domain adaptive filter," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 38, no. 2, pp. 373–376, 1990.
[11] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*, Springer-Verlag, Berlin, 2001.
[12] H. V. Sorensen, D. L. Jones, M. T. Heideman, and C. S. Burrus, "Real-valued fast Fourier transform algorithms," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 35, no. 6, pp. 849–863, 1987.
[13] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1996.
[14] T. Gänsler and J. Benesty, "A frequency-domain double-talk detector based on a normalized cross-correlation vector," *Signal Processing*, vol. 81, no. 8, pp. 1783–1787, 2001.
[15] P. J. Huber, *Robust Statistics*, Wiley, New York, NY, USA, 1981.



Jacob Benesty was born in 1963. He received his M.S. degree in microwaves from Pierre & Marie Curie University, France, in 1987, and his Ph.D. degree in control and signal processing from Orsay University, France, in April 1991. During his Ph.D. (from November 1989 to April 1991), he worked on adaptive filters and fast algorithms at the Centre National d'Etudes des Télécommunications (CNET), Paris, France. From January 1994 to July 1995, he worked at Telecom Paris University on multichannel adaptive filters and acoustic echo cancellation. He joined Bell Labs, Lucent Technologies (formerly AT&T) in October 1995, first as a Consultant and then as a Member of Technical Staff. Since this date, he has been working on stereophonic acoustic echo cancellation, adaptive algorithms, source localization, robust network echo cancellation, and blind identification. He was the Cochair of the 1999 International Workshop on Acoustic Echo and Noise Control. He coauthored *Advances in Network and Acoustic Echo Cancellation* (Springer-Verlag, Berlin, 2001). He is also a coeditor/coauthor of *Acoustic Signal Processing for Telecommunication* (Kluwer Academic Publishers, Boston, 2000) and *Adaptive Signal Processing: Applications to Real-World Problems* (Springer-Verlag, Berlin, 2003).

Tomas Gänsler was born in Sweden in 1966. He received his M.S. degree in electrical engineering and his Ph.D. degree in signal processing from Lund University, Lund, Sweden, in 1990 and 1996. From 1997 to September 1999, he held a position as an Assistant Professor at Lund University. During 1998, he was employed by Bell Labs, Lucent Technologies as a Consultant and from October 1999, he became a Member of Technical Staff. Since 2001, he has been with Agere Systems, a spin-off from Lucent Technologies' Microelectronics Group. His research interests include robust estimation, adaptive filtering, mono/multichannel echo cancellation, and subband signal processing. He coauthored *Advances in Network and Acoustic Echo Cancellation* and he is also a coauthor of *Acoustic Signal Processing for Telecommunication*.