

Convergence Analysis of a Frequency-Domain Adaptive Filter with Exponential Power Averaging and Generalized Window Function

PIET C. W. SOMMEN, PIET J. VAN GERWEN, SENIOR MEMBER, IEEE,
HENK J. KOTMANS, AND A. J. E. M. JANSSEN

Abstract—One of the advantages of a Frequency-Domain Adaptive Filter (FDAF) is that one can achieve convergence at a constant rate over the whole frequency range by choosing the adaptation constant for each frequency bin l equal to the overall adaptation constant divided by an estimate of the input power at this frequency bin. A commonly used method, applied in this paper, to estimate the input power is to do an exponentially weighting with smoothing constant β on the magnitude squared of the input values at each frequency bin l . Furthermore, it is known that a correctly implemented FDAF, using the overlap-save method, contains five $2N$ -points Fast Fourier Transforms (FFT). Two of these are used to force the last N points of the time-domain augmented impulse response to zero by applying a particular window function. In this paper, an analysis is given of the FDAF where the window function is generalized. Using these results, the convergence behavior of FDAF's with various window functions is compared. Furthermore, the analysis describes the influence of β on the convergence behavior of the FDAF over the whole convergence range.

I. INTRODUCTION

ADAPTIVE DIGITAL filters are extremely useful devices in many applications of digital signal processing, including channel equalization, sensor array processing, and echo and noise interference cancellation. In this paper, we will restrict ourselves to an echo cancellation structure for acoustic applications. Typical examples of this kind of applications are the loudspeaking telephone [1] and audio teleconferencing [2] of which the basic echo cancellation scheme is given in Fig. 1. The speech signal $x(k)$ from the "far end" speaker reflects via an acoustic echo path as an echo signal $e(k)$. This acoustic path can be considered as a multireflection medium with an impulse response which may have lengths up to several hundreds of milliseconds. Together with the "near end" signal $s(k)$, this echo $e(k)$ arrives at a microphone. The adaptive filter uses a model of the acoustic echo path and makes a replica $\hat{e}(k)$ of the echo signal $e(k)$. Thus, the adaptive filter cancels the echoes of speech signal $x(k)$ on the "forward path" which appear on the "return path." Theoretically the residual signal $r(k) = s(k) + e(k) - \hat{e}(k)$ in steady state will almost be equal to the signal $s(k)$.

The two main problems with adaptive transversal filters for acoustic echo cancellation configurations are

- the number of weights N , needed for the adaptive filter to model an acoustic path, is very large, viz., from 500 up to 2000;
- the input signal $x(k)$ is a correlated signal.

These two difficulties can be tackled with a Frequency-Domain Adaptive Filter (FDAF). Namely:

- Using an FDAF results in block processing in which one block of input data is processed simultaneously, producing one block of output data. This block processing can be done by efficient algorithms such as Fast Fourier Transforms (FFT). In this way, the amount of computational requirements in terms of multiply-adds per one block of N output samples can be greatly reduced compared with time-domain approaches. This is accomplished by replacing convolution with a multiplication of transforms which implies a complexity reduction from $O(N^2)$ to $O(N \log(N))$.
- The eigenvalues of the input autocorrelation matrix are given approximately by uniformly spaced samples of the input power spectrum. This implies that weights associated with frequencies having little power converge more slowly than those associated with frequencies having greater power. A large variation in the input power spectrum leads to highly disparate eigenvalues and therefore highly disparate time constants, some of which may be very large. Frequency-domain techniques can easily be modified to allow more uniform convergence of the weights of the adaptive process. The weights are adapted independently from each other and this corresponds approximately to one-tap Least Mean Square (LMS) adaptive filters. The time constant of the l th weight, assuming stationary inputs, is inversely proportional to αP_x , where α is the adaptation constant and P_x is the input power related to that weight. In order to make all weights converge at the same rate, the adaptation constant can be made different for each

Manuscript received August 28, 1986; revised March 3, 1987.
The authors are with Philips Research Laboratories, 5600 JA Eindhoven, The Netherlands.
IEEE Log Number 8714713.

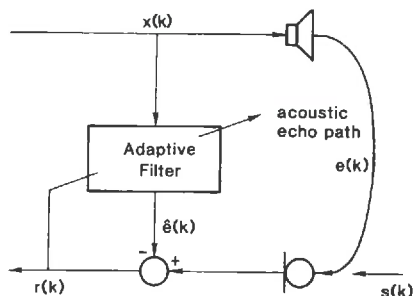


Fig. 1. Basic echo cancellation scheme for acoustic applications.

frequency bin according to $\mu_l = \alpha / \hat{P}_{X_l}$, where \hat{P}_{X_l} is an estimate of the input power at the l th weight.

A disadvantage of an FDAF is that the linear convolution must be accomplished by a circular one. This can be done by using the overlap-save method which implies that FFT's are needed of length $2N$. An overlap-save implemented FDAF as described by Clark *et al.* [3] requires five FFT's. Two of them are used to force the last N points of the time-domain augmented weights to zero. This zero forcing has been done by using a particular window function in the time domain. Mansour [4] proposes an FDAF without a window function. This configuration is less complex (three FFT's), but the convergence behavior is worse than the FDAF as proposed by Clark [3].

In this paper, an analysis is given of the FDAF where the window function is generalized. Using these results, the convergence behavior of FDAF's with different window functions is compared. From this it can be shown that there are efficient window functions with complexity almost equal to the FDAF as proposed by Mansour [4], while the convergence behavior is comparable to the configuration proposed by Clark [3].

Since in practical applications we do not know the power spectrum of the input signal $x(k)$, this is normally calculated by some averaging process. For this reason, the analysis given in this paper will furthermore describe the convergence behavior of the FDAF when the estimation of the input power P_{X_l} is done by using an exponentially weighted average, with smoothing constant β , of the magnitude squared of the input values at frequency bin l . From this part, it appears that for a small final misadjustment, which implies a small adaptation constant α as used in data transmission, we can vary the rate of convergence over a very large range by choosing different β . On the other hand, when dealing with a large final misadjustment, and thus with large α , the influence of β on the rate of convergence is very small. In all cases, however, the best choice for β is as large as possible, which is in contrast to the choice $2\alpha = 1 - \beta$ as suggested in [5, p. 174].

To support the theory, some graphs will give theoretical and simulation results describing the convergence behavior of the FDAF. As references to work done in the same field, we mention [6] which gives an analysis of the FDAF where the power estimation is done by uniformly averaging over the last K measurements. The paper [6] does not describe the influence of the window function and the

overlapping of the input sequence. In [7] and [9], an analysis is given of a Time Domain Block LMS algorithm while [8] and [9] give an analysis of the Time Domain LMS algorithm. In [10] the influence of the smoothing constant β is given for small adaptation constant α .

II. OVERLAP-SAVE IMPLEMENTATION OF AN FDAF

It is well known that linear convolution/correlation in the time domain may be performed by multiplication in the frequency domain. This can very easily be implemented with the aid of FFT's by using the overlap-save method. Fig. 2 shows for the echo cancellation problem the overlap-save implemented FDAF as described by Clark *et al.* [3]. In this figure, the FFT's are denoted by F's, while signal paths with double lines in the figures refer to paths in the frequency domain, and single lines refer to time-domain signal paths. In the text, we will use lower case characters for the time-domain signals, while upper case characters are used for frequency-domain signals. The unknown echo path impulse response is given by h' , while the length of this echo path equals N . Denoting the time-domain adaptive weights by $w_l(m)$, the adaptive filter has to perform a linear convolution between the input signal $x(k)$ and these weights. To do this, the input signal is segmented into blocks of length $2N$. These blocks are transformed to the frequency domain by a $2N$ points FFT. The l th frequency bin $X_l(m)$ in the m th data block is multiplied by the weight $W_l(m)$, which is the Fourier transform of the N adaptive weights $w_l(m)$ augmented with N zeros, to obtain the filter output $\hat{E}_l(m)$. The overlap-save procedure is now executed by overlapping the input segments over a length of N (segm1), discarding the first N points of the circular convolution output (segm2) and choosing a window function g as

$$g_i = \begin{cases} 1 & \text{for } i = 0, \dots, N-1 \\ 0 & \text{for } i = N, \dots, 2N-1 \end{cases} \quad (1)$$

to force the last N points of the weight function w to zero. The principle of the frequency-domain LMS algorithm is to update the weights as long as there is correlation between the signals $x(k)$ and $r(k)$. The overlap-save procedure to determine this correlation is now executed by overlapping the input segments over a length of N (segm1), discarding the last N points of the circular output, which is implicitly done by the time-domain windowing function g_i as defined in (1), and augmenting with N zeros in front of the segment of $r(k)$ (segm3).

Denoting matrices by bold-face characters and vectors by underlined bold-face characters, the frequency-domain algorithm in vector-matrix notation becomes

$$\underline{W}(m+1) = \underline{W}(m) + 2\alpha \underline{F} g \underline{F}^{-1} \hat{P}_X^{-1}(m) X^*(m) \underline{R}'(m)$$

$$\underline{W}(0) = (0, \dots, 0)^T \quad \text{with } m = 0, 1, \dots \quad (2)$$

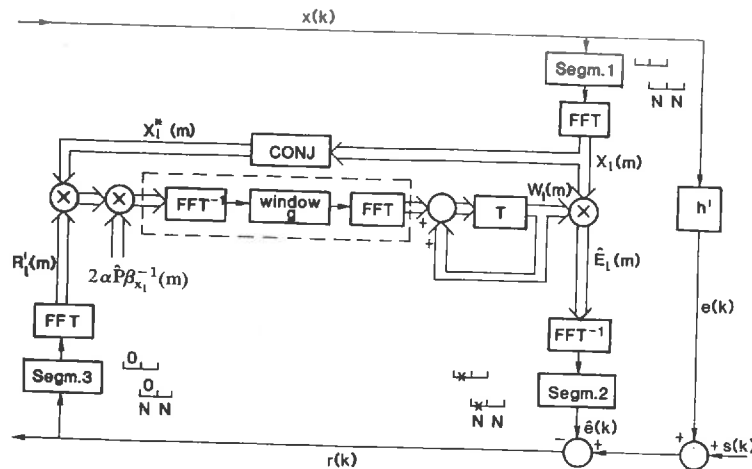


Fig. 2. Overlap-save implementation of FDAF.

where

$$\underline{W}(m) = (W_0(m), \dots, W_{2N-1}(m))^T$$

α

F

$$g = \text{diag}(1, \dots, 1, 0, \dots, 0)$$

$$\hat{P}_X(m) = \text{diag}(\hat{P}_{X_0}(m), \dots, \hat{P}_{X_{2N-1}}(m))$$

$$X(m) = \text{diag}(X_0(m), \dots, X_l(m), \dots, X_{2N-1}(m))$$

$$X^*(m)$$

$$X_l(m) = \sum_{i=0}^{2N-1} x((m-1)N+i) e^{-j\theta i}$$

$$\underline{R}'(m) = (R'_0(m), \dots, R'_{2N-1}(m))^T$$

weight vector; T denotes transpose,

adaptation constant,

$2N$ -point FFT matrix,

diagonal window matrix defined in (1),

estimate of diagonal input power matrix,

input signal matrix in frequency domain,

complex conjugate transpose of $X(m)$,

FFT of $x(k)$, $\theta = \pi/N$, $x(k) = 0$ for $k < 0$,

residual signal vector in frequency domain.

From (2), it follows that the weights $W_l(m)$ are updated as long as there is correlation between the signals $x(k)$ and $r(k)$. This correlation is calculated by the product $X^*(m)\underline{R}'(m)$. By decorrelating the input signal, the convergence speed can be accelerated [4]. This can be accomplished by normalizing the input power spectrum which is done by the inverse of the estimate of the input power matrix $\hat{P}_X^{-1}(m)$. The factor FgF^{-1} achieves the windowing.

III. ANALYSIS OF AN FDAF WITH EXPONENTIAL POWER AVERAGING AND GENERALIZED WINDOW FUNCTION

To analyze the FDAF, we first give some definitions, notations, and assumptions in Section III-A, while in Section III-B, the most important characteristics are given to describe the dynamic behavior of the adaptive filter. In Section III-C, the analysis is given of the dynamic behavior of the FDAF (2), where the window function g is generalized, while the power estimation is done with an exponential averaging network. Namely

$$\hat{P}_{X_l}(m) = \beta \hat{P}_{X_l}(m-1) + (1-\beta) |X_l(m)|^2 \quad (3)$$

with $0 < \beta < 1$

where β is the smoothing constant of the power averaging network.

Since all processing is done with block processing techniques, the description of the FDAF is carried out in vector-matrix notation. For this reason, we refer to Fig. 3 for the analysis, which is an equivalent of Fig. 2, where all signals are in vector-matrix notation.

A. Definitions, Notations, and Assumptions

We assume the frequency bins of the input signal to consist of independent complex Gaussian stationary random variables with zero mean. This implies that

$$E[X_k^*(m)X_l(m)] = \begin{cases} 0 & \text{for } k \neq l \\ P_{X_l} & \text{for } k = l \end{cases} \quad (4)$$

where $E[\cdot]$ is the mathematical expectation. This independence assumption implies that we assume the input signal to have an autocorrelation function over maximal N points. All signals are segmented into blocks and these blocks are described by vectors. The input signal in the frequency domain, however, is represented in a diagonal matrix since this notation allows the adaptive algorithm (2) to be described with well-known vector-matrix arithmetics. Furthermore, we assume that each frequency bin

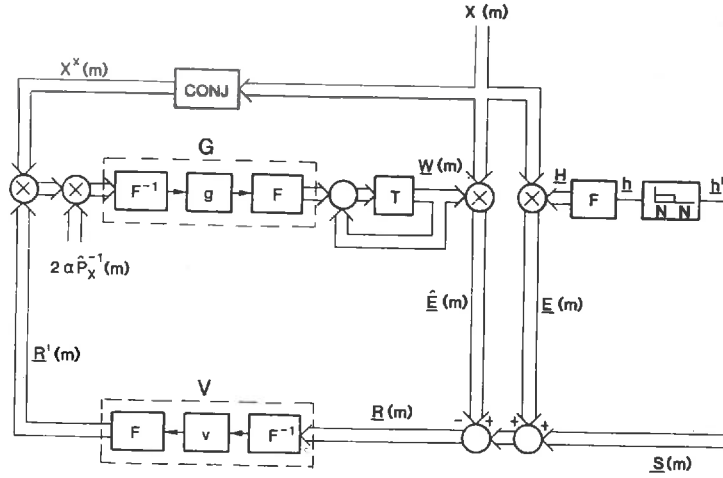


Fig. 3. Vector-matrix diagram of FDAF.

$X_k(0), X_k(1), \dots$, is jointly Gaussian distributed with covariance matrix

$$\begin{pmatrix} 1 & -\frac{1}{2} & & \\ -\frac{1}{2} & 1 & -\frac{1}{2} & \\ & -\frac{1}{2} & 1 & -\frac{1}{2} \\ & & & \ddots \end{pmatrix} \cdot P_{X_k}. \quad (5)$$

When the input signal is a white-noise signal, this covariance matrix is exact because of the 50-percent overlapping of the input sequence by N samples (segm1 from Fig. 2). From experiments, it appears that input signals which can be modeled by an all-pole filter have almost the same covariance matrix.

The power matrix of the input signal is diagonal and is defined by

$$P_X = E[X^*(m)X(m)] = \text{diag}(P_{X_0}, \dots, P_{X_{2N-1}}). \quad (6)$$

Using the same approximations as given in [11], where the circular autocorrelation matrix C_x is constructed from the Toeplitz autocorrelation matrix R_x in such a way that their eigenvalues are approximately the same, we have

$$P_X = F \cdot C_x \cdot F^*. \quad (7)$$

The frequency bins $S_i(m)$ of the signal $s(k)$ are also assumed to be complex Gaussian stationary random variables with zero mean. These frequency bins are defined as

$$S_i(m) = \sum_{i=0}^{2N-1} s((m-1)N+i) e^{-j\theta i l} \quad (8)$$

with $s(k) = 0$ for $k < 0$;

Furthermore, it is assumed that $X_l(m)$ and $S_l(m)$ are independent. The purpose of the acoustic canceller is to cancel the echoes of speech on the "forward path" which appear on the "return path." Conversations always will contain periods during which speech is present in both directions at the same time (double talk). Since this double

talk problem is beyond the scope of this paper, we will assume that a double-talk detector is incorporated to prevent misadjustment of the echo canceller. For this reason, the signal $s(k)$ may be represented by a white-noise signal which includes all imperfections. This implies that the power spectrum of the signal $s(k)$ is flat, i.e.,

$$E[|S_l(m)|^2] = P_S \quad \forall l$$

and

$$P_S = \text{diag}(P_{S_0}, \dots, P_{S_{2N-1}}) = P_S \cdot I \quad (9)$$

where I is the $2N \times 2N$ identity matrix. From Parseval's relation, it follows that the average power of one block in the frequency domain is equal to the total power of one block in the time domain. This implies for signal $s(k)$

$$\begin{aligned} \frac{1}{2N} \sum_{l=0}^{2N-1} P_{S_l} &= \frac{1}{2N} \cdot 2N \cdot P_S = P_S \\ &= \sum_{i=0}^{2N-1} E[s^2((m-1)N+i)]. \end{aligned} \quad (10)$$

The window matrix g is generalized as

$$g = \text{diag}(g_0, \dots, g_{2N-1}) \quad \text{with } 0 \leq g_i \leq 1. \quad (11)$$

The segmentation blocks (segm2) and (segm3) of Fig. 2 may be combined as a segmentation window v , which is defined as

$$v_i = \begin{cases} 0 & \text{for } i = 0, \dots, N-1 \\ 1 & \text{for } i = N, \dots, 2N-1. \end{cases} \quad (12)$$

In Fig. 3, this function is generalized as the segmentation matrix

$$v = \text{diag}(v_0, \dots, v_{2N-1}) \quad \text{with } 0 \leq v_i \leq 1. \quad (13)$$

Since g and v are diagonal, the matrices

$$G = FgF^{-1} \quad \text{and} \quad V = FvF^{-1} \quad (14)$$

are circulant [12]. In Appendix I, we use the fact that most of the energy of the matrix $C_v = V \cdot V^*$ is concentrated at the diagonals $(C_v)_{l,l+1}$, $(C_v)_{l,l}$, and $(C_v)_{l,l-1}$ which are the main and first codiagonals of C_v . This implies that the generalization (13) must be within this scope.

The Fourier transforms are carried out with the $2N$ -point FFT matrix F from which the (k, l) th element is given by

$$(F)_{kl} = e^{-j\theta kl} \quad \text{with } \theta = \pi/N \quad (15)$$

and with the property

$$F^{-1} = \frac{1}{2N} \cdot F^* \quad (16)$$

The variance of the difference signal $\tilde{e}(k) = e(k) - \hat{e}(k)$ appears to be an important quantity to describe the dynamic behavior of the adaptive filter. The frequency-domain transformation from the m th block of this difference signal is

$$\begin{aligned} \underline{\tilde{E}} &= \underline{E}(m) - \underline{\hat{E}}(m) = X(m) \cdot \underline{H} - X(m) \cdot \underline{W}(m) \\ &= X(m) \cdot \underline{D}(m) \end{aligned} \quad (17)$$

where

$$\begin{aligned} \underline{H} &= F \cdot \underline{h} \\ \underline{h} &= ((\underline{h}')^T, 0, \dots, 0)^T \\ \underline{h}' &= (h_0, \dots, h_{N-1})^T \\ \underline{W}(m) &= F \cdot \underline{w}(m) \\ \underline{w}(m) &= (w_0(m), \dots, w_{2N-1}(m))^T \\ \underline{D}(m) &= \underline{H} - \underline{W}(m) = F \cdot \underline{d}(m) \\ \underline{d}(m) &= \underline{h} - \underline{w}(m) \end{aligned}$$

The Block Mean Square Error (BMSE), which is equal to the average power of one block of the difference signal in the frequency domain, is given by

$$\begin{aligned} P_{\tilde{E}}(m) &= \frac{1}{2N} E \{ \{ X(m) \cdot \underline{D}(m) \}^* \cdot \{ X(m) \cdot \underline{D}(m) \} \} \\ &= \frac{1}{2N} \text{trace} (P_X \cdot \Delta(m)) \end{aligned} \quad (18)$$

where the assumption is made that $X(m)$ and $\underline{D}(m)$ are independent while $\Delta(m)$ is the covariance matrix of the difference vector defined by

$$\Delta(m) = E \{ \underline{D}(m) \underline{D}^*(m) \} \quad (19)$$

while $\text{trace}(\cdot)$ is the trace of a matrix. This covariance matrix is related to the time-domain covariance matrix as

$$\Delta(m) = F \delta(m) F^*$$

with

$$\delta(m) = E \{ \underline{d}(m) \underline{d}^T(m) \} \quad (20)$$

Another quantity we will need is the minimum attainable BMSE, which is equal to the average power of one block of the echo signal in the frequency domain. This is defined as

$$\begin{aligned} P_E &= \frac{1}{2N} E \{ \{ X(m) \underline{H} \}^* \cdot \{ X(m) \underline{H} \} \} \\ &= \sum_{i=0}^{2N-1} E \{ e^{2((m-1)N+i)} \} \end{aligned} \quad (21)$$

B. Convergence Behavior Characteristics

The characteristics of interest which describe the convergence behavior of the FDAF are as follows.

Stability: The largest value α_0 of the adaptation constant α that yields a stable algorithm. This implies

$$\lim_{m \rightarrow \infty} P_{\tilde{E}}(m) = P_{\tilde{E}} \quad \text{exists if and only if } 0 < \alpha < \alpha_0. \quad (22)$$

Final Misadjustment: The fractional amount by which the steady-state BMSE exceeds the minimum attainable BMSE is called the final misadjustment and is defined by

$$M = \lim_{m \rightarrow \infty} \left(\frac{P_{\tilde{E}}(m)}{P_E} \right) = \frac{P_{\tilde{E}}}{P_E} \quad (23)$$

The Rate of Convergence: The rate of convergence of the BMSE to its steady state can be measured by the sum of

frequency-domain impulse response vector,
augmented time-domain impulse response vector,
original impulse response vector of length N ,
frequency-domain weight vector,
time-domain weight vector,
frequency-domain difference vector,
time-domain difference vector.

the following series [7], [8]:

$$J = \sum_{m=0}^{\infty} [P_{\tilde{E}}(m) - P_{\tilde{E}}] \quad (24)$$

with small J indicating fast convergence. From (24), it is clear that J is the "total area" under the function $P_{\tilde{E}}(m) - P_{\tilde{E}}$. Fitting through $P_{\tilde{E}}(m)$ an exponential function, with time constant α , defined as

$$P_{\tilde{E}}(m) = (P_{\tilde{E}}(0) - P_{\tilde{E}}) e^{-2m/\tau} + P_{\tilde{E}} \quad (25)$$

and expressing J as

$$J = f(\alpha, \beta) \cdot (P_{\tilde{E}}(0) - P_{\tilde{E}}) \quad (26)$$

$$\tau = \frac{-2}{\ln [1 - (1/f(\alpha, \beta))]} \quad (27)$$

is evident.

Another well-known quantity characterizing the rate of convergence is v_{20} [13], which gives the number of iterations which are required to reduce the residual signal $r(k)$ by 20 dB. The relations between v_{20} and τ is

$$v_{20} = \frac{10}{10 \log(e)} \cdot \tau \approx 2.3 \tau \quad (28)$$

C. Analysis

With the notations as introduced in Section III-A, the update algorithm (2) becomes

$$\underline{W}(m+1) = \underline{W}(m) + 2\alpha \underline{G} \hat{P}_X^{-1}(m) X^*(m) \underline{V} \underline{R}(m) \quad (29)$$

By using (see Fig. 3)

$$\underline{R}(m) = X(m) \cdot \underline{D}(m) + \underline{S}(m) \quad (30)$$

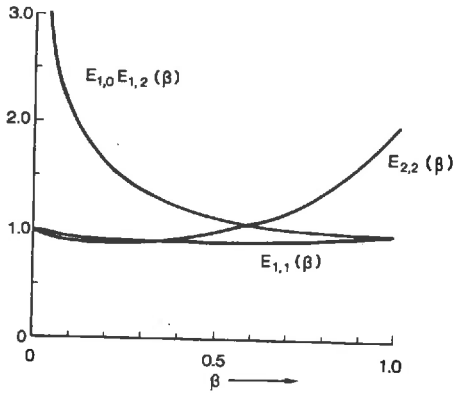


Fig. 4. The factors $E_{ij}(\beta) = E[|X_k(m)|^{2i} \hat{P}_{X_k}^{-j}(m)]$ as a function of β .

the update algorithm (29) becomes

$$\underline{D}(m+1) = \{I - 2\alpha \mathbf{G} \hat{P}_X^{-1}(m) X^*(m) V X(m)\} \cdot \underline{D}(m) - 2\alpha \mathbf{G} \hat{P}_X^{-1}(m) X^*(m) V \underline{S}(m). \quad (31)$$

We will first set up the difference equation for $\Delta(m)$. By using the independency assumption between the matrix $X(m)$ and the (zero mean) vector $\underline{S}(m)$, this difference equation follows from (31) and is given by

$$\begin{aligned} \Delta(m+1) = & \Delta(m) - 2\alpha \mathbf{G} \cdot E[\hat{P}_X^{-1}(m) X^*(m) V X(m)] \cdot \Delta(m) \\ & - 2\alpha \Delta(m) \cdot E[X^*(m) V^* X(m) \hat{P}_X^{-1}(m)] \cdot \mathbf{G}^* \\ & + 4\alpha^2 \mathbf{G} \cdot E[\hat{P}_X^{-1}(m) X^*(m) V X(m) \cdot \Delta(m) \cdot X^*(m) V^* X(m) \hat{P}_X^{-1}(m)] \cdot \mathbf{G}^* \\ & + 4\alpha^2 \mathbf{G} \cdot E[\hat{P}_X^{-1}(m) X^*(m) V \cdot E[\underline{S}(m) \underline{S}^*(m)]] \cdot V^* X(m) \hat{P}_X^{-1}(m) \cdot \mathbf{G}^*. \end{aligned} \quad (32)$$

Using the results of Appendix I, we get

$$\begin{aligned} \Delta(m+1) \approx & \Delta(m) - 2\alpha E_{11}(\beta) \bar{\Sigma}_v \mathbf{G} \cdot \Delta(m) \\ & - 2\alpha E_{11}(\beta) \bar{\Sigma}_v \Delta(m) \cdot \mathbf{G}^* \\ & + 4\alpha^2 \mathbf{G} \cdot \{E_{11}^2(\beta) \bar{\Sigma}_v^2 \Delta(m) \\ & + E_v(\beta) \text{diag}(\underline{\Delta}(m))\} \cdot \mathbf{G}^* \\ & + 4\alpha^2 P_S E_{12} E_{10}(\beta) \bar{\Sigma}_{v^2} \mathbf{G} \cdot P_X^{-1} \cdot \mathbf{G}^*. \end{aligned} \quad (33)$$

The vector $\underline{\Delta}(m)$ contains the diagonal elements of the matrix $\Delta(m)$, while the "average area" functions $\bar{\Sigma}_v$ and $\bar{\Sigma}_{v^2}$ are given by

$$\bar{\Sigma}_v = \frac{1}{2N} \sum_{i=0}^{2N-1} v_i \quad \text{and} \quad \bar{\Sigma}_{v^2} = \frac{1}{2N} \sum_{i=0}^{2N-1} v_i^2. \quad (34)$$

The function $E_{ij}(\beta)$ is defined as the mathematical expectation

$$E_{ij}(\beta) = E[|X_k(m)|^{2i} \hat{P}_{X_k}^{-j}(m)] \quad (35)$$

where $\hat{P}_{X_k}(m)$ is the exponentially weighted power average as defined in (3) with smoothing constant β . For our analysis, we need among other things $E_{11}(\beta)$ and $E_{22}(\beta)$, which are dimensionless quantities and therefore independent of the frequency bin k . Furthermore, the product $E_{12}(\beta)E_{10}(\beta)$ is needed, which is also dimensionless and thus independent of k . For convenience, this product will be abbreviated as $E_{12}E_{10}(\beta)$. Explicit formulas for $E_{ij}(\beta)$

are given in Appendix II, while plots of $E_{11}(\beta)$, $E_{22}(\beta)$, and $E_{12}E_{10}(\beta)$ are given in Fig. 4. Furthermore, $E_v(\beta)$ in (33) is composed from the functions $E_{ij}(\beta)$ according to

$$E_v(\beta) = \{E_{22}(\beta) - E_{11}^2(\beta) - E_{12}E_{10}(\beta)\} \cdot \bar{\Sigma}_v^2 + E_{12}E_{10}(\beta) \cdot \bar{\Sigma}_{v^2}. \quad (36)$$

Using (20), the time-domain transformation of (33) is given by

$$\begin{aligned} \delta(m+1) \approx & \delta(m) - 2\alpha E_{11}(\beta) \bar{\Sigma}_v \mathbf{g} \cdot \delta(m) \\ & - 2\alpha E_{11}(\beta) \bar{\Sigma}_v \delta(m) \cdot \mathbf{g} \\ & + 4\alpha^2 \mathbf{g} \cdot \{E_{11}^2(\beta) \bar{\Sigma}_v^2 \delta(m) + E_v(\beta) \\ & \cdot [F^{-1} \text{diag}(\underline{\Delta}(m))(F^{-1})^*]\} \cdot \mathbf{g} \\ & + 4\alpha^2 P_S E_{12} E_{10}(\beta) \bar{\Sigma}_{v^2} \mathbf{g} \cdot [F^{-1} P_X^{-1}(F^{-1})^*] \cdot \mathbf{g}. \end{aligned} \quad (37)$$

Since $(\delta(m))_{kl}^2 \leq (\delta(m))_{kk} \cdot (\delta(m))_{ll}$, the convergence to zero of the diagonal elements of $\delta(m)$ ensures the convergence of the off-diagonal elements. We shall therefore concentrate on the dynamic behavior of the diagonal elements of the matrix $\delta(m)$. Some time after the conver-

gence process has started, the matrix $\delta(m)$ may be approximated by a diagonal matrix. This implies that $\Delta(m)$, which is related to $\delta(m)$ by (20), is a circulant matrix [12].

By defining $\underline{\delta}(m)$ as the vector containing the diagonal elements of the matrix $\delta(m)$ and using (7) we can rewrite (37) as

$$\underline{\delta}(m+1) \approx A \cdot \underline{\delta}(m) + b P_S \bar{P}'_X g^2 \cdot \underline{1}, \quad m = 0, 1, \dots \quad (38)$$

where

$$A = \left\{ \left(I - 2\alpha E_{11}(\beta) \bar{\Sigma}_v \mathbf{g} \right)^2 + 4\alpha^2 \frac{1}{2N} E_v(\beta) g^2 \cdot \underline{1} \cdot \underline{1}^T \right\}$$

$$b = 4\alpha^2 \frac{1}{2N} E_{12} E_{10}(\beta) \bar{\Sigma}_{v^2}$$

$$\underline{1} = (1, \dots, 1)^T$$

$$\bar{P}'_X = \frac{1}{2N} (\mathbf{R}_X^{-1})_{00}$$

1) *Stability*: The largest value of the adaptation constant α_0 that yields a stable algorithm depends on the behavior of the vector $\underline{\delta}(m)$. The convergence of this vector depends on the eigenvalues of matrix A . Convergence occurs if and only if the eigenvalues of A are all within the unit circle. Similar to the derivation given in [8], we get

$$\alpha > 0$$

and

$$1 + \frac{4\alpha^2}{2N} E_v(\beta) \sum_{i=0}^{2N-1} \frac{g_i^2}{\{1 - 2\alpha E_{11}(\beta) \bar{\Sigma}_v g_i\}^2 - 1} > 0 \quad (39)$$

which implies that

$$\alpha_0 > 0 \quad \text{and} \quad \alpha_0 \frac{E_v(\beta)}{E_{11}(\beta) \bar{\Sigma}_v} \bar{\Sigma}_{g, \alpha_0} < 1 \quad (40)$$

where $\bar{\Sigma}_{g, \alpha}$ is a "weighted area" function defined as

$$\bar{\Sigma}_{g, \alpha} = \frac{1}{2N} \sum_{i=0}^{2N-1} \frac{g_i}{1 - \alpha E_{11}(\beta) \bar{\Sigma}_v g_i} \quad (41)$$

2) *Final Misadjustment*: Since $\delta(m)$ is diagonal in final state, we can rewrite $P_{\bar{E}}$ from (18) using (7) as

$$P_{\bar{E}}(m) = \bar{P}_X \text{trace}(\delta(m)) = \bar{P}_X \mathbf{1}^T \cdot \delta(m) \quad (42)$$

with the average power \bar{P}_X defined as

$$\bar{P}_X = 2N(\mathbf{R}_x)_{00} \quad (43)$$

The final misadjustment is now given by

$$\begin{aligned} M &= P_{S/E} \cdot \tilde{P}_X \mathbf{b} \mathbf{1}^T \cdot (\mathbf{I} - \mathbf{A})^{-1} \cdot \mathbf{g}^2 \cdot \mathbf{1} \\ &= P_{S/E} \cdot \epsilon \end{aligned} \quad (44)$$

with

$$P_{S/E} = P_S/P_E \quad \text{and} \quad \tilde{P}_X = (\mathbf{R}_x)_{00} \cdot (\mathbf{R}_x^{-1})_{00} \quad (45)$$

For an input signal with a flat spectrum, the factor \tilde{P}_X equals one. When the input signal is taken to be a highly correlated signal for which the spectrum is given by $1 - \cos(\phi)$, this factor is two. In general, this factor will be close to one. For this reason, we will assume for simplicity that $\tilde{P}_X \approx 1$. Using the Bartlett formula as given in [8], we can write

$$\mathbf{1}^T \cdot (\mathbf{I} - \mathbf{A})^{-1} \cdot \mathbf{g}^2 = (\dots, f_i(\alpha, \beta), \dots) \quad (46)$$

with

$$f_i(\alpha, \beta) = \left(\frac{g_i}{4\alpha E_{11}(\beta) \bar{\Sigma}_v (1 - \alpha E_{11}(\beta) \bar{\Sigma}_v g_i)} \right) \left/ \left(1 - \alpha \frac{E_v(\beta)}{E_{11}(\beta)} \frac{1}{\bar{\Sigma}_v} \bar{\Sigma}_{g, \alpha} \right) \right. \quad (47)$$

With this, the expression for the final misadjustment becomes

$$M = P_{S/E} \left(\alpha \frac{E_{12} E_{10}(\beta)}{E_{11}(\beta)} \frac{\bar{\Sigma}_v^2}{\bar{\Sigma}_v} \bar{\Sigma}_{g, \alpha} \right) \left/ \left(1 - \alpha \frac{E_v(\beta)}{E_{11}(\beta)} \frac{1}{\bar{\Sigma}_v} \bar{\Sigma}_{g, \alpha} \right) \right. \quad (48)$$

3) *Rate of Convergence*: The rate of convergence J is given by

$$J = \bar{P}_X \mathbf{1}^T \sum_{m=0}^{\infty} (\delta(m) - \delta) = \bar{P}_X \mathbf{1}^T \cdot (\mathbf{I} - \mathbf{A})^{-1} \cdot (\delta(0) - \delta) \quad (49)$$

where $\delta = \lim_{m \rightarrow \infty} \delta(m)$. Using the results of the Bartlett

formula (47), J can be rewritten as

$$J = \bar{P}_X \sum_{i=0}^{2N-1} f_i(\alpha, \beta) \cdot ((\delta(0))_i - (\delta)_i) \quad (50)$$

where the functions $f_i(\alpha, \beta)$ are given by (47). The time constant of the i th weight is, similar to (27), given by

$$\tau_i = \frac{-2}{\ln[1 - (1/f_i(\alpha, \beta))]} \quad (51)$$

The overall time constant τ of the adaptive filter is a function of all τ_i , but is mainly determined by those τ_i of the weights w_i which have to converge to the largest value h_i of the echo path impulse response. In general, we can use the *a priori* knowledge that the absolute value of the global envelope of the echo path impulse response is a decreasing function with i , while we assume for simplicity at this moment that $|h_i|$ is maximal at $i=0$. For this reason, the overall time constant can be approximated by $\tau \approx \tau_0$.

IV. ANALYSIS AND SIMULATION RESULTS FOR THREE WINDOW FUNCTIONS

In this section, we will study the convergence behavior of FDAF's with various window functions and an exponential power averaging with smoothing constant β both by using the simulation and analytical results as derived in the foregoing section. Window functions of interest are as follows.

- Clark [3] proposes an overlap-save FDAF configuration which contains five FFT's with a window function as defined in (1). The used window will be referred to as the "Block- N " window.

- Mansour [4] proposes an FDAF configuration without a window function which contains three FFT's. In Fig. 3, this implies a "shortcut" window function which will be referred to as the "Block- $2N$ " window and is defined as $g_i = 1$ for $i = 0, \dots, 2N-1$.

- Using the *a priori* knowledge about the decreasing behavior of the global envelope of the echo path impulse response, an efficient window function was proposed in [14] for an FDAF configuration containing three FFT's. This window function will be referred to as the "Cosine" window and is defined as $g_i = \frac{1}{2} + \frac{1}{2} \cos(\theta i)$ for $i = 0, \dots, 2N-1$ with $\theta = \pi/N$.

Although the segmentation window v is generalized in the analysis, we will assume here that it is defined as in (12). This is the segmentation window as it appears in the overlap-save configuration of Fig. 2. This implies that $\bar{\Sigma}_v = \bar{\Sigma}_{v^2} = \frac{1}{2}$ and $\bar{\Sigma}_v^2 = \frac{1}{4}$. Furthermore, Table I summarizes the analytical results, describing the convergence behavior of the FDAF's both as a function of the three mentioned window functions and the smoothing constant β .

The final misadjustment M is expressed in decibels as $10 \log(M) = 10 \log(P_{S/E}) + 10 \log(\epsilon)$, with ϵ defined as in (44). The quantity $10 \log(P_{S/E}) = 10 \log(P_S) - 10 \log(P_E)$ gives the ratio of the power level of (noise) signal $s(k)$ to

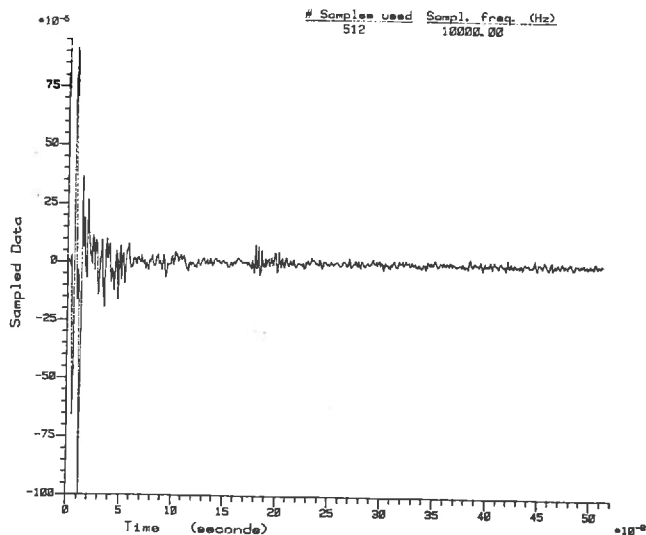


Fig. 5. Impulse response of loudspeaking intercom system.

TABLE I
CONVERGENCE BEHAVIOR OF FDAF WITH THREE PARTICULAR WINDOW FUNCTIONS

	Stability $0 < \alpha < \alpha_0$	Final Misadjustment $M = P_S/E \cdot \epsilon$	Rate of Coverage $\nu_{20} \approx 2.3 \frac{-2}{\ln[1-1/(1-\alpha\beta)]}$
Block-N	$\alpha_{0,N} = \frac{4}{E_1 + E_2 + E_3}$	$\epsilon_N = E_3 \cdot \frac{1}{2} \cdot \frac{1 - \frac{\alpha}{\alpha_{0,N}}}{1 - \frac{\alpha}{\alpha_{0,N}}}$	$\nu_{20,N} = \frac{-4.6}{\ln[1 - 2\alpha E_1 (1 - \frac{\alpha}{\alpha_{0,N}})]}$
Block-2N	$\alpha_{0,2N} = \frac{2}{E_2 + E_3}$	$\epsilon_{2N} = E_3 \cdot \frac{1 - \frac{\alpha}{\alpha_{0,2N}}}{1 - \frac{\alpha}{\alpha_{0,2N}}}$	$\nu_{20,2N} = \frac{-4.6}{\ln[1 - 2\alpha E_1 (1 - \frac{\alpha}{\alpha_{0,2N}})]}$
Cosine	$\alpha_{0,C} = 2 \cdot \frac{(2E_2 + 2E_3 - E_1)}{(E_2 + E_3)^2}$	$\epsilon_C = E_3 \cdot \frac{1}{2} \cdot \frac{1 - \frac{\alpha}{\alpha_{0,C}}}{1 - \frac{\alpha}{\alpha_{0,C}}}$	$\nu_{20,C} \approx \frac{-4.6}{\ln[1 - 2\alpha E_1 (1 - \frac{\alpha}{\alpha_{0,C}}) g_2(\alpha)]}$
$g_1(\alpha) = \frac{E_2 + E_3}{4} + \frac{E_1}{4(1 + \sqrt{1 - \alpha \frac{E_1}{E_2}})}$ $g_2(\alpha) = \frac{E_2 + E_3}{E_1} - \frac{E_2 + E_3 - E_1}{E_1 \sqrt{1 - \alpha \frac{E_1}{E_2}}}$			
$E_1 = E_{11}(\beta) \quad E_2 = \frac{E_{12}(\beta)}{E_{11}(\beta)} \quad E_3 = \frac{E_{13}E_{10}(\beta)}{E_{11}(\beta)}$			

the power level of the echo signal $e(k)$ in one block. In our simulations, this factor was equal to -30 dB. For the echo path, we choose the impulse response of a loudspeaking intercom system from which the impulse response (Fig. 5) was sampled at 10 kHz which results in $N = 512$. The FFT's are thus of length $2N = 1024$. To simulate with "speech like" signals, a white-noise signal was passed through a formant filter (12th-order all-pole filter).

Both final misadjustment and rate of convergence are functions of the adaptation constant α . In this paper, we are not interested in the actual value of α (this value can be calculated by the formulas given in Table I), but in the convergence properties as a function of both the smoothing constant β and of various window functions. For this reason, we made curves (Figs. 6 and 7) in which α was eliminated by construction. These curves show the final misadjustment on the vertical axis and the rate of convergence on the horizontal axis with α varying in the range $0 < \alpha < \alpha_0$. The $\alpha = 0$ point is at the lower right corner, where a very good final misadjustment is reached after a very long time, while the $\alpha = \alpha_0$ point is in the upper right corner, where it takes a long time to reach a very bad final misadjustment. Since the upper part of the curve gives the same rate of convergence as the lower part, but with a

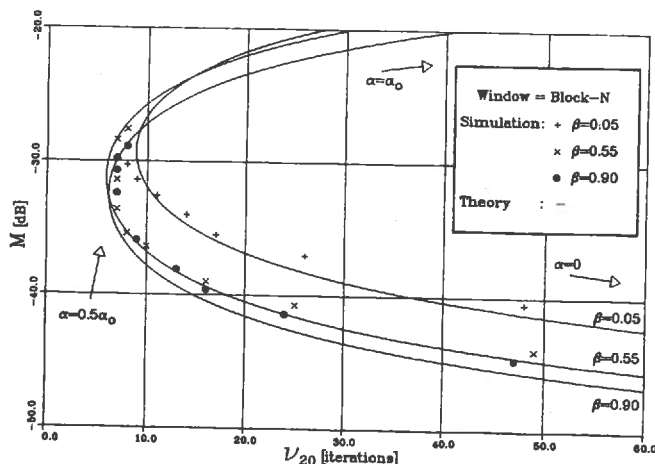


Fig. 6. Convergence behavior of FDAF as a function of β with Block-N window function.

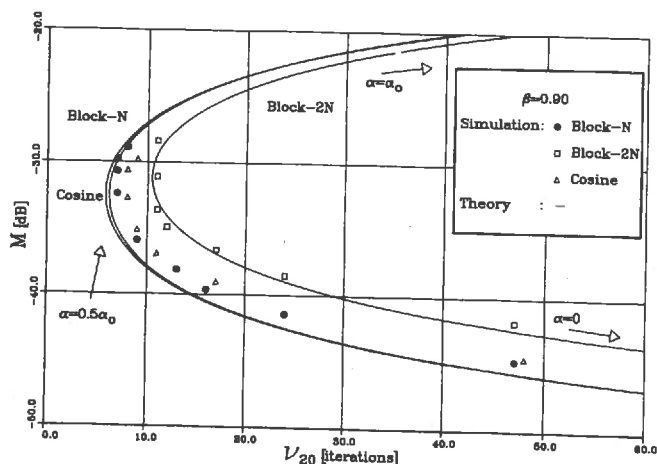


Fig. 7. Convergence behavior of FDAF for three particular window functions, with $\beta = 0.9$.

worse final misadjustment, it is clear that only the lower part for $0 < \alpha < \alpha_0/2$ is of practical interest. Fig. 6 shows a curve which gives the final misadjustment $10 \log(M)$ as a function of the rate of convergence ν_{20} with the smoothing factor β as a parameter. The bounds for the smoothing constant are $0 < \beta < 1$, while three different curves for $\beta = 0.05, 0.55, \text{ and } 0.9$ are plotted in Fig. 6. In general, β should not be chosen too close to the bounds $\beta \approx 0$ or $\beta \approx 1$. For $\beta \approx 0$, the exponential network does not average. This implies that, for each l th frequency bin, the adaptation constant α is divided by its momentary value of the input power spectrum. This value may become very small and can cause the algorithm to become unstable. Also, we should not take β too close to 1 because then the convergence behavior is very much dependent on the initial values of $\hat{P}_{X_i}(0)$. For that case, we have $\hat{P}_{X_i}(m) \approx \hat{P}_{X_i}(0) \forall m$. Since we did not describe this initialization effect in our analysis (for this we refer to [10]), we initialized \hat{P}_X with the power spectrum of the input signal.

From simulations and analytical results, it appears that the influence of the smoothing constant β on the convergence behavior of the FDAF for all window functions is similar. For that reason, Fig. 6 only gives the results for

the Block- N window. Fig. 7 depicts the results of the effect from the various window functions on the convergence behavior of the FDAF with $\beta = 0.9$. From Fig. 5, we see that the echo path impulse response h_i is nonzero for i in the neighborhood of $N-1$. To overcome the problem that the Cosine window would have a relatively too small value in comparison to h_i for $i \approx N-1$, we have shifted the Cosine window over some distance. Namely, $g_i = \frac{1}{2} + \frac{1}{2} \cos(\theta(i-40))$ for $i = 0, \dots, 2N-1$.

V. CONCLUSIONS

From Fig. 6, the following conclusions can be reached.

1) For small final misadjustment, and hence a very small adaptation constant α , the influence of β is very large. In that case, we can vary the rate of convergence over a very large range by choosing a different β . This is in agreement with the result of [10], where the analysis was made for a very small α . For α near to $\alpha_0/2$, which are the points in the curves with the smallest v_{20} , the influence of β is negligible.

2) The number of iterations used to reach a final misadjustment decreases when β increases. This implies that we should take the smoothing constant β as large as possible, taking into account the initialization effect as mentioned before. This in contrast to the choice $2\alpha = 1 - \beta$ as suggested in [5, p. 174].

3) The analytical results are systematically a little too low. The reason for this is the approximation made for the factor $10 \log(\hat{P}_X) = 10 \log((R_x)_{00} \cdot (R_x^{-1})_{00}) \approx 0$ dB. For a nonflat spectrum, this factor may rise up to 2 or 3 dB.

From Fig. 7, the following conclusions can be reached.

4) The convergence properties of the FDAF with the Cosine window are comparable to the configuration with the Block- N window. The FDAF with the Cosine window, however, can be implemented very efficiently in the frequency domain, as suggested in [14], and contains only three FFT's, whereas the Block- N configuration contains five FFT's.

5) The number of iterations needed to reach a certain final misadjustment for the FDAF with the Block-2 N window, realized with three FFT's, is about twice as high as for the configuration with the Block- N or Cosine window. In practical situations, the echo path impulse response may have some delay τ_d . The Cosine window can be shifted, without significant complexity increase, while still having convergence properties which are comparable to the Block- N configuration. In general, we can say that *a priori* knowledge about the global envelope of the echo path impulse response can be used to reduce complexity.

As a final remark, we mention that the analysis of the FDAF with nonstationary input signals and the analysis of the tracking capabilities of the FDAF are beyond the scope of this paper but are subjects for future research.

APPENDIX I

ANALYSIS OF THE EXPECTATION MATRICES IN (32)

In this appendix, we give the mathematical expectations of the matrices which appear in (32). For simplicity we define

$$E_a = E[\hat{P}_X^{-1}(m) X^*(m) V X(m)] \quad (A1)$$

$$E_b = E[\hat{P}_X^{-1}(m) X^*(m) V X(m) \cdot \Delta(m) \cdot X^*(m) V^* X(m) \hat{P}_X^{-1}(m)] \quad (A2)$$

$$E_c = E[\hat{P}_X^{-1}(m) X^*(m) V \cdot E[\underline{S}(m) \underline{S}^*(m)] \cdot V^* X(m) \hat{P}_X^{-1}(m)]. \quad (A3)$$

Using the Gaussian assumption and by denoting the (k, l) th element of matrix by $(\cdot)_{kl}$ and using the fact that V is circulant, we get

$$(E_a)_{kl} = E[\hat{P}_{X_k}^{-1}(m) X_k^*(m) X_l(m)] \cdot (V)_{kl} = E_{11}(\beta) \cdot \bar{\Sigma}_v \quad \text{for } k = l \\ = 0 \quad \text{for } k \neq l. \quad (A4)$$

The (k, l) th element of matrix E_b is given by

$$(E_b)_{kl} = \sum_q \sum_t E[\hat{P}_{X_k}^{-1}(m) \hat{P}_{X_l}^{-1}(m) X_k^*(m) X_t(m) X_q^*(m) X_l(m)] \cdot (V)_{kt} (V^*)_{qt} \cdot (\Delta(m))_{tq}. \quad (A5)$$

We will evaluate (A5) for $k = l$ and for $k \neq l$. For $k = l$

$$(E_b)_{kk} = \sum_q \sum_t E[\hat{P}_{X_k}^{-2}(m) |X_k(m)|^2 X_q^*(m) X_t(m)] \cdot (V)_{kt} (V^*)_{qk} \cdot (\Delta(m))_{tq} \\ = \sum_t E[\hat{P}_{X_k}^{-2}(m) |X_k(m)|^2 |X_t(m)|^2] \cdot |(V)_{kt}|^2 \cdot (\Delta(m))_{tt} \\ = E_{22}(\beta) \bar{\Sigma}_v^2 (\Delta(m))_{kk} + E[\hat{P}_{X_k}^{-2}(m) |X_k(m)|^2] \sum_{t \neq k} E[|X_t(m)|^2] |(V)_{kt}|^2 (\Delta(m))_{tt} \\ = E_{22}(\beta) \bar{\Sigma}_v^2 (\Delta(m))_{kk} - E_{12} E_{10}(\beta) \bar{\Sigma}_v^2 (\Delta(m))_{kk} \\ + E[\hat{P}_{X_k}^{-2}(m) |X_k(m)|^2] \sum_t E[|X_t(m)|^2] (\Delta(m))_{tt} |(V)_{kt}|^2. \quad (A6)$$

For the moment, we will concentrate on the last term of this equation. Assuming that $|(V)_{kl}|^2$ is mainly concentrated on the diagonals $(\cdot)_{k,k+1}$, $(\cdot)_{k,k}$, and $(\cdot)_{k,k-1}$ and making the assumption that

$$E[|X_{k+1}(m)|^2] \cdot E[|D_{k+1}(m)|^2] \approx E[|X_k(m)|^2] \cdot E[|D_k(m)|^2] \approx E[|X_{k-1}(m)|^2] \cdot E[|D_{k-1}(m)|^2] \quad (A7)$$

we can write this last term as

$$\begin{aligned} E[\hat{P}_{X_k}^{-2}(m)|X_k(m)|^2] \sum_t E[|X_t(m)|^2](\Delta(m))_{tt} |(V)_{kt}|^2 \\ \approx E[\hat{P}_{X_k}^{-2}(m)|X_k(m)|^2] \\ \cdot E[|X_k(m)|^2] \cdot (\Delta(m))_{kk} \sum_t |(V)_{kt}|^2 \\ \approx E_{12} E_{10}(\beta) \cdot (\Delta(m))_{kk} \cdot \bar{\Sigma}_{v^2}. \end{aligned} \quad (A8)$$

This results in

$$\begin{aligned} (E_b)_{kk} \approx (E_{22}(\beta) - E_{12} E_{10}(\beta)) \bar{\Sigma}_v^2 (\Delta(m))_{kk} \\ + E_{12} E_{10}(\beta) \bar{\Sigma}_{v^2} (\Delta(m))_{kk}. \end{aligned} \quad (A9)$$

The nondiagonal elements of E_b are

$$(E_b)_{kl} = E_{11}^2(\beta) \bar{\Sigma}_v^2 (\Delta(m))_{kl}. \quad (A10)$$

Combining the equations for $k=l$ and $k \neq l$ gives

$$\begin{aligned} (E_b)_{kl} = E_{11}^2(\beta) \bar{\Sigma}_v^2 (\Delta(m))_{kl} \\ + E_v(\beta) (\Delta(m))_{kk} \quad \text{for } k=l \\ = E_{11}^2(\beta) \bar{\Sigma}_v^2 (\Delta(m))_{kl} \quad \text{for } k \neq l. \end{aligned} \quad (A11)$$

$$E_{ij}(\beta) = \lim_{m \rightarrow \infty} E_{ij}(\beta; m) = \frac{\pi^{j-i}!}{(j-i)!(1-\beta)^j} \int_0^\infty \left(\frac{1-G(\lambda)}{\lambda} \right)^i \frac{\lambda^{j-1}}{F(\lambda)} d\lambda \quad i=0,1,\dots, j=1,2,\dots$$

$$E_{i0}(\beta) = \pi^{-i}!, \quad i=0,1,\dots$$

In this equation, $E_v(\beta)$ is defined as

$$\begin{aligned} E_v(\beta) = (E_{22}(\beta) - E_{11}^2(\beta) - E_{12} E_{10}(\beta)) \bar{\Sigma}_v^2 \\ + E_{12} E_{10}(\beta) \bar{\Sigma}_{v^2}. \end{aligned} \quad (A12)$$

For the last matrix, we have

$$\begin{aligned} (E_c)_{kl} = P_S \sum_t E[\hat{P}_{X_k}^{-1}(m) \hat{P}_{X_l}^{-1}(m) X_k^*(m) X_l(m)] \\ \cdot (V)_{kt} (V^*)_{tl} \\ = P_S E_{12} E_{10}(\beta) \bar{\Sigma}_{v^2} P_{X_k}^{-1} \quad \text{for } k=l. \end{aligned} \quad (A13)$$

APPENDIX II

ANALYSIS OF THE NUMBERS $E_{ij}(\beta)$ IN (35)

We reproduce in this appendix the results of [15 and 16] as far as relevant for the present paper. We only need to consider

$$E_{11}(\beta), E_{22}(\beta), E_{12} E_{10}(\beta) = E_{12}(\beta) \cdot E_{10}(\beta). \quad (A14)$$

From the definition in (35) and the assumption in Section III-A on the joint probability density of $X_k(0), X_k(1), \dots$, it follows that the quantities in (A14) are independent of $E_{10}(\beta) = P_{X_k}$. It is convenient to choose $E_{10}(\beta) = 1/\pi$. Note also that we now can drop the index k in the

right-hand side of (35). The results we reproduce are for the limiting case $m \rightarrow \infty$, which is realistic since m in (35) is usually large, but at the end of this appendix we shall indicate how to deal with the case of finite m .

It is slightly more convenient to replace m by $m-1$ in (35). We have

$$\begin{aligned} E_{ij}(\beta; m) = E[\hat{P}_X^{-j}(m-1) \cdot |X(m-1)|^{2i}] \\ = \frac{1}{|Q_m|} \int_{C^m} \frac{|z_0|^{2i} e^{(-\pi(Q_m^{-1}z, z))}}{(1-\beta) \sum_{l=0}^{m-1} \beta^l |z_l|^2)^j} dz \end{aligned} \quad (A15)$$

where we have set $\mathbf{z} = (z_0, \dots, z_{m-1})^T = (X(m-1), \dots, X(0))^T$. Here, Q_m is the leading $m \times m$ section of the infinite matrix Q , given by

$$Q = \begin{pmatrix} 1 & -\frac{1}{2} & & & \\ -\frac{1}{2} & 1 & -\frac{1}{2} & & \\ & -\frac{1}{2} & 1 & -\frac{1}{2} & \\ & & & \ddots & \ddots \\ & & & & \ddots & \ddots \end{pmatrix} \quad (A16)$$

and $|Q_m|$ denotes the determinant of Q_m . In [16], it is shown that

Here, $G(\lambda) = ((I + \lambda U)^{-1})_{1,1}$, the left upper corner element of $(I + \lambda U)^{-1}$, $F(\lambda) = |I + \lambda U|$, the determinant of $I + \lambda U$, and U is the infinite matrix

$$U = \begin{pmatrix} 1 & -\frac{1}{2}\beta^{1/2} & & & \\ -\frac{1}{2}\beta^{1/2} & \beta & -\frac{1}{2}\beta^{3/2} & & \\ & -\frac{1}{2}\beta^{3/2} & \beta^2 & -\frac{1}{2}\beta^{5/2} & \\ & & & \ddots & \ddots \\ & & & & \ddots & \ddots \end{pmatrix}. \quad (A18)$$

The functions $G(\lambda), F(\lambda), \lambda \geq 0$ can be calculated conveniently according to the formulas

$$G(\lambda) = \frac{1}{1 + \lambda - \frac{1}{4}\beta\lambda^2 G(\lambda\beta)} \quad \text{and} \quad \frac{1}{F(\lambda)} = \prod_{k=0}^{\infty} G(\lambda\beta^k). \quad (A19)$$

Formula (A19) allows for a very rapidly converging continuous fraction expansion of $G(\lambda)$. Moreover, in [16, sec. 5] it is shown that $1/F(\lambda)$ decays rapidly (like $e^{(s^2/2\gamma)}$) with $s = \log(\lambda/2\beta^{1/2}), \gamma = \log(\beta)$, especially when β is close to 1. Hence, it is feasible to calculate the integrals in (A17) numerically. Details are presented in [16, sec. 2].

Some of the inequalities derived in [16, sec. 4.1] are

$$E_{11}(\beta) \leq E_{12}E_{10}(\beta) \quad \frac{1}{2-\beta} \leq E_{11}(\beta) \leq 1$$

$$E_{22}(\beta) \leq \frac{2}{2-\beta}. \quad (\text{A20})$$

Furthermore, it is shown in [16, sec. 4.2, sec. 4.3] that $E_{12}E_{10}(\beta) \rightarrow \infty$ as $\beta \downarrow 0$, that $E_{22}(\beta) < E_{11}(\beta)$ for β close to 0, and that $E_{12}E_{10}(\beta) < 1$ for β close to 1, and the limiting behavior of the quantities in (A14) has been determined. Fig. 4 gives the plots of these quantities as a function of $0 < \beta < 1$.

We note that many of the results given here also hold for the case of finite m . One just has to replace $G(\lambda)$ and $F(\lambda)$ in (A19) by $G_m(\lambda) = ((I + \lambda U)_m^{-1})_{1,1}$ and $F_m(\lambda) = \|(I + \lambda U)_m\|$, respectively, where the index m refers to taking the leading $m \times m$ section of $I + \lambda U$. The relevant formulas for this case are given in [15, sec. 2] and [16, eq. (2.19)].

ACKNOWLEDGMENT

The authors wish to thank T. A. C. M. Claasen for his encouragement to submit this paper for this special issue. Furthermore, he and J. M. M. Verbakel are acknowledged for the stimulating discussions and helpful suggestions. B. van Welt is acknowledged for carrying out the simulations.

REFERENCES

- [1] C. R. South, C. E. Hoppitt, and A. V. Lewis, "Adaptive filters to improve a loudspeaking telephone," *Electron. Lett.*, vol. 15, no. 21, pp. 673-674, Oct. 1979.
- [2] O. A. Horna, "Echo control in teleconferencing," in *Proc. GLOBECOM'83* (San Diego), pp. 16.2.1-7.
- [3] G. A. Clark, S. R. Parker, and S. K. Mitra, "A unified approach to time- and frequency-domain realization for FIR adaptive digital filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, no. 5, pp. 1073-1083, Oct. 1983.
- [4] D. Mansour and A. H. Gray, Jr., "Unconstrained frequency-domain adaptive filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-30, no. 5, pp. 726-734, Oct. 1982.
- [5] C. F. N. Cowan and P. M. Grant, *Adaptive Filters*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [6] N. J. Bershad and P. L. Feintuch, "A normalized frequency domain LMS adaptive algorithm," submitted to *IEEE ASSP* 1986.
- [7] A. Feuer, "Performance analysis of the block least mean square algorithm," *IEEE Trans. Circuits Syst.*, vol. CAS-32, pp. 960-963, Sept. 1985.
- [8] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 1, pp. 222-230, Feb. 1985.
- [9] W. A. Gardner, "Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique," *Signal Process.* no. 6, pp. 113-133, 1984.
- [10] J. C. Lee and C. K. Un, "Performance of transform-domain LMS adaptive digital filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 3, pp. 499-509, June 1986.
- [11] R. M. Gray, "Toeplitz and circulant matrices: A Review," Stanford Electronics Laboratories, Tech. Rep. 6502-1, 1971.
- [12] P. J. Davis, *Circulant Matrices*. New York: Wiley, 1979.
- [13] N. A. M. Verhoeckx, H. C. van den Elzen, W. A. M. Sniijders, and P. J. van Gerwen, "Digital echo cancellation for baseband data transmission," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, no. 6, pp. 768-781, Dec. 1979.
- [14] P. C. W. Sommen, "Frequency domain filter with an efficient window function," in *Proc. ICC'86* (Toronto, Canada), June 1986, pp. 60.6.1-5.
- [15] A. J. E. M. Janssen, "On the eigenvalues of an infinite Jacobi matrix," *Philips J. Research*, vol. 40, pp. 323-351, 1985.
- [16] A. J. E. M. Janssen, "On certain integrals occurring in the analysis of a frequency domain, power compensated adaptive filter," submitted to *Philips J. Research*.



Piet C. W. Sommen was born in Ulicoten, The Netherlands, on February 17, 1954. He received the Ingenieur degree in electrical engineering from the Technological University, Delft, The Netherlands, in 1981.

In 1981, he joined the Philips Research Laboratories, Eindhoven, The Netherlands, where he was engaged in research on CAD for circuit design. Since 1984, his field of interest is adaptive digital signal processing.

✱



Piet J. van Gerwen (SM'77) was born in Eindhoven, The Netherlands, on March 30, 1930. He received the degree in electrical engineering from the Eindhoven School of Technology, Eindhoven, in 1952.

He joined the Philips Research Laboratories, Eindhoven, in 1954 and completed the training program in physics in 1957. He was then engaged in research on systems for speech bandwidth reduction and companders. At present, he is working on the transmission of digital signals

with special emphasis on coding methods, data transmission, echo cancellation, and digital signal processing.

✱



Henk J. Kotmans was born in Eindhoven, The Netherlands, on December 10, 1952. He received the degree in electrical engineering from the Polytechnical Highschool, Eindhoven.

In 1976, he joined the Philips Research Laboratories, Eindhoven, The Netherlands. From 1976 to 1982, he was engaged in the field of speech encoding and recognition. His present work involves adaptive digital signal processing for data transmission and acoustic applications.

✱



A. J. E. M. Janssen was born in Breda, The Netherlands, in 1953. He received the Engineer degree and the Ph.D. degree in mathematics in 1976 and 1979, respectively, both from the Eindhoven University of Technology, Eindhoven, The Netherlands.

From 1979 to 1981, he was with the California Institute of Technology, Pasadena, as a Bateman Research Instructor in Mathematics. He is presently with the Philips Research Laboratories, Eindhoven, The Netherlands. His interests are

time-frequency descriptions of signals with emphasis on fundamental aspects, spectral estimation, and applications of mathematical analysis in signal theory.