

STATE-SPACE ARCHITECTURE OF THE PARTITIONED-BLOCK-BASED ACOUSTIC ECHO CONTROLLER

Fabian Kuech, Edwin Mabande

Fraunhofer IIS
Am Wolfsmantel 33
D-91058 Erlangen, Germany
fabian.kuech@iis.fraunhofer.de

Gerald Enzner

Institute of Communication Acoustics
Ruhr-Universität Bochum
D-44780 Bochum, Germany
gerald.enzner@rub.de

ABSTRACT

Acoustic echo cancellation has traditionally employed basically all variants known from deterministic adaptive filter design, such as least mean-square (LMS), recursive least-squares (RLS), and frequency-domain adaptive filters (FDAF). More recently, a stochastic adaptive filter design based on the concept of acoustic state-space modeling of the echo path has been introduced to accommodate for an ever sought unification of adaptive filtering and adaptation control. The corresponding Kalman filter theory has been formulated for single-channel, multi-channel, and nonlinear echo cancellation problems. This paper closes an important gap by formulating the state-space model and the corresponding adaptive algorithm for the partitioned-block filtering structure which is especially relevant in practice. This structure allows for the use of significantly longer filter lengths in comparison to previous work, and for the flexible design and implementation of acoustic echo cancellers for widely differing acoustic conditions.

Index Terms— acoustic echo control, adaptive filtering

1. INTRODUCTION AND RELATION TO PRIOR WORK

Acoustic echo represents a well-known distraction in hands-free voice communication systems. Due to acoustic coupling between the loudspeaker and the microphone of a telecommunication terminal, the far-end talker receives a delayed version of his own voice that will eventually inhibit fluent conversation. The general setup of the acoustic echo problem is illustrated in Fig. 1. The far-end signal $x(n)$ is played back by the near-end loudspeaker. The microphone signal $y(n)$ then picks up the echo $d(n)$ together with the near-end signal $s(n)$, including background noise and local speech. The adaptive acoustic echo canceler (AEC) regenerates and subtracts an estimate of the echo from the microphone signal. A variety of adaptive filter structures and methods to control the adaptation in adverse environments have been proposed for this purpose [1, 2, 3]. Typically, due to time-varying acoustics and echo-path undermodeling, the AEC is not always able to sufficiently remove the echo and, thus, a residual echo suppressor (RES) is introduced after the AEC

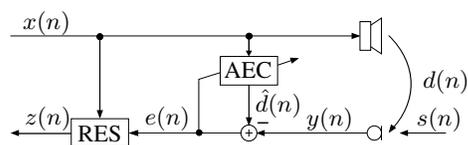


Fig. 1. Setup of the acoustic echo cancellation problem.

to attenuate remaining echo components [4, 5, 6, 7].

In modern communication systems such as Voice-over-IP services or high-quality video conferencing systems, sampling rates of 16 kHz and higher are used. This implies a significant increase in computational complexity for the AEC. Moreover, the convergence speed of time-domain adaptive filters is usually not sufficient in case of high sampling rates and long echo paths. Besides subband adaptive filters [8, 9], frequency-domain adaptive filters using block processing are well-known solutions to address both of these issues [10, 11, 12, 13]. However, the required length of the frequency transform becomes relatively large for long echo paths, leading to potential algorithmic noise, e.g., when implementing the AEC with fixed point arithmetic in embedded devices. In this case, approaches based on partitioned-block filtering [14, 15, 16, 17, 18] are more suitable, as they allow for flexible designs, e.g., a separate choice of the transformation length and the time span covered by the AEC, and the filter length is not bounded to powers of two as typically used in fast fourier transform implementations. Moreover, a reduction of the transformation length also reduces the algorithmic delay caused by post-processing stages such as the RES. By using a partitioned block structure, approximations of the RES filter to reduce the delay as, e.g., proposed in [19], can be avoided.

The step-size parameter to control the adaptation of the filter coefficients is generally a critical component of the AEC. An overview of popular methods is found, e.g., in [1, 2, 3]. The alternative approach in [6] relies on an acoustic state-space model of the echo path to deduce a robust and efficient frequency-domain adaptive filter with inherent step-size control, however, not considering block partitioning. This approach has analogously been applied to multi-channel as well as nonlinear adaptive filtering problems [20, 21]. In this paper, we extend the approach in [6] to the partitioned-block filtering structure, which is of practical importance for the application of AECs for widely varying acoustic conditions.

In our paper, Sec. 2 revises partitioned-block adaptive filtering based on state-space modeling in analogy with [6]. Sec. 3 shows that the corresponding exact Kalman filter in the block-frequency-domain can be simplified into a diagonalized version. This light approximation leads to a variant of the known multi-delay adaptive filter [14], but providing inherent adaptive step-size control. While the resulting step-size turns out to be similar to the one in [15], we propose a different estimator for the required system distance, again in analogy with the non-partitioned algorithm in [6]. In Sec. 4, a so far unique relation between the step-size parameters of all partitions and the optimum RES filter is derived in order to efficiently obtain the RES coefficients. Simulation results in Sec. 5 confirm the suitability of the proposed architecture for acoustic echo control.

2. ACOUSTIC ECHO CANCELLATION (AEC) PART

At first, the previous concept of acoustic state-space modeling is formally extended to partitioned-block filters. On this basis, the corresponding Kalman filter for acoustic echo path tracking is formulated.

2.1. State-Space Partitioned-Block Echo Path Model

Using the notation in Fig. 1, the microphone signal $y(n)$ can be expressed as the sum of the near-end signal $s(n)$ and the echo signal $d(n)$. Since the echo signal results from the discrete-time convolution of the loudspeaker input signal $x(n)$ with the acoustic echo path $w(n)$, we have

$$y(n) = x(n) * w(n) + s(n). \quad (1)$$

In the following, we assume that the acoustic echo path can be sufficiently modeled by a corresponding finite impulse response (FIR) filter. Aiming at a partitioned block implementation of (1), we divide the FIR filter with coefficients $w(n)$ into B partitions of length L . The coefficient vector $\mathbf{w}_b(k)$ of length L then contains the coefficients of the b -th partition:

$$\mathbf{w}_b(k) = [w(bL, k), w(bL + 1, k), \dots, w(bL + L - 1, k)]^T. \quad (2)$$

We further introduce an input vector $\mathbf{x}_b(k)$ of the b -th partition of length $M > L$ for the block time index k and a frame shift of R :

$$\mathbf{x}_b(k) = [x(kR - bL - M + 1), \dots, x(kR - bL)]^T \quad (3)$$

Then, the corresponding complex valued excitation matrix $\mathbf{X}_b(k)$ in the frequency domain is obtained as

$$\mathbf{X}_b(k) = \text{diag}\{\mathbf{F}_M \mathbf{x}_b(k)\}, \quad (4)$$

where \mathbf{F}_M is the Fourier matrix of size $M \times M$. Here, $\text{diag}\{\mathbf{a}\}$ denotes a diagonal matrix with the vector \mathbf{a} on its main diagonal. The elements $X_b(m, k)$ on the main diagonal of $\mathbf{X}_b(k)$ are given by

$$\mathbf{F}_M \mathbf{x}_b(k) = [X_b(0, k), X_b(1, k), \dots, X_b(M - 1, k)]^T, \quad (5)$$

where m denotes the frequency index. The frequency domain representation $\mathbf{W}_b(k)$ of the filter partitions is given by

$$\mathbf{W}_b(k) = [W_b(0, k), W_b(1, k), \dots, W_b(M - 1, k)]^T, \quad (6)$$

where the constraint is applied that only the first L coefficients of the time-domain correspondence are non-zero:

$$\mathbf{W}_b(k) = \mathbf{F}_M \begin{bmatrix} \mathbf{w}_b(k) \\ \mathbf{0} \end{bmatrix}. \quad (7)$$

Applying the overlap-and-save method for computing a block of microphone signal [22], we have

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{y}(k) \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{s}(k) \end{bmatrix} + \mathbf{Q}_V \mathbf{F}_M^{-1} \sum_{b=0}^{B-1} \mathbf{X}_b(k) \mathbf{W}_b(k), \quad (8)$$

where \mathbf{Q}_V denotes the windowing matrix

$$\mathbf{Q}_V = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_V \end{bmatrix} \quad (9)$$

and \mathbf{I}_V is the identity matrix of size $V \times V$. The signal vectors $\mathbf{y}(k)$ and $\mathbf{s}(k)$ in (8) then contain the $V = M - L + 1$ latest samples of the microphone and the near-end signal, respectively:

$$\mathbf{y}(k) = [y(kR - V + 1), y(kR - V + 2), \dots, y(kR)]^T \quad (10)$$

$$\mathbf{s}(k) = [s(kR - V + 1), s(kR - V + 2), \dots, s(kR)]^T \quad (11)$$

Note that V represents the number of valid samples of $y(k)$, which result from fast convolution of the loudspeaker signals and the echo path. Note that $V > R$ is well possible for specific choices of the DFT length M and the partition size L . In this case, $\mathbf{y}(k)$ contains only R new samples, whereas $V - R$ other valid samples have already been computed in the previous frame [14].

The frequency domain version of (8) is obtained by left multiplying it with the Fourier matrix \mathbf{F}_M :

$$\begin{aligned} \mathbf{Y}(k) &= \mathbf{S}(k) + \mathbf{F}_M \mathbf{Q}_V \mathbf{F}_M^{-1} \sum_{b=0}^{B-1} \mathbf{X}_b(k) \mathbf{W}_b(k) \\ &= \mathbf{S}(k) + \sum_{b=0}^{B-1} \mathbf{C}_b(k) \mathbf{W}_b(k), \end{aligned} \quad (12)$$

where $\mathbf{C}_b(k) = \mathbf{F}_M \mathbf{Q}_V \mathbf{F}_M^{-1} \mathbf{X}_b(k)$ has been introduced.

In the following we define variables which will be useful for the presentation of the Kalman filter. Let the coefficients of the adaptive filter partitions in the frequency domain be denoted by $\hat{\mathbf{W}}_b(k)$. Then, the frequency-domain coefficient error vector for the b -th partition is defined as

$$\mathbf{W}_{b,r}(k) = \mathbf{W}_b(k) - \hat{\mathbf{W}}_b(k). \quad (13)$$

The corresponding covariance matrix is given by

$$\mathbf{P}_b(k) = \mathcal{E}\{\mathbf{W}_{b,r}(k) \mathbf{W}_{b,r}^H(k)\}, \quad (14)$$

where $\mathcal{E}\{\}$ denotes the expectation operator. Regarding (12) and (13), the frequency-domain error $\mathbf{E}(k)$ at the output of the echo canceler can be expressed in terms of the coefficient error vector $\mathbf{W}_{b,r}(k)$, i.e.,

$$\mathbf{E}(k) = \mathbf{Y}(k) - \sum_{b=0}^{B-1} \mathbf{C}_b(k) \hat{\mathbf{W}}_b(k) \quad (15)$$

$$= \sum_{b=0}^{B-1} \mathbf{C}_b(k) \mathbf{W}_{b,r}(k) + \mathbf{S}(k). \quad (16)$$

In general, the acoustic echo path, and, thus, the coefficients $\mathbf{W}_b(k)$, are assumed to be slowly time-varying. Following [6], we introduce a simple statistical Markov model for the dynamic behavior of the filter partitions $\mathbf{W}_b(k)$ according to

$$\mathbf{W}_b(k + 1) = \mathbf{W}_b(k) + \Delta \mathbf{W}_b(k). \quad (17)$$

Analogously to [6], the desired stochastic state-space model for the filter partitions is then defined by the Markov model in (17) together with the linear observation model in (12).

2.2. Exact Kalman Filter Solution

Regarding the derivation of the Kalman filter for the partitioned-block model with state-space architecture, i.e., Eqs. (17) and (12), we can rely on a derivation presented in [20] for the problem of multi-channel adaptive filtering based on acoustic state-space modeling. Here, it is important to note that the partitioned-block implementation of a linear filter according to (12) can be interpreted as a specific multiple-input single-output system, where the input signal of the b -th channel (b -th partition) is given by $\mathbf{X}_b(k)$. Using [20, 21], the equations describing the partitioned-block version of the Kalman filter are then given for the b -th partition as

$$\hat{\mathbf{W}}_b(k + 1) = A \hat{\mathbf{W}}_b^+(k), \quad (18)$$

$$\mathbf{P}_b(k + 1) = A^2 \mathbf{P}_b^+(k) + \Psi_{b,\Delta\Delta}(k), \quad (19)$$

$$\hat{\mathbf{W}}_b^+(k) = \hat{\mathbf{W}}_b(k) + \mathbf{K}_b(k) \mathbf{E}(k), \quad (20)$$

$$\mathbf{P}_b^+(k) = [\mathbf{I}_M - \mathbf{K}_b(k) \mathbf{C}_b(k)] \mathbf{P}_b(k), \quad (21)$$

with the so-called Kalman gain of the b -th partition,

$$\mathbf{K}_b(k) = \mathbf{P}_b(k) \mathbf{C}_b^H(k) \left[\sum_{b=0}^{B-1} \mathbf{C}_b(k) \mathbf{P}_b(k) \mathbf{C}_b^H(k) + \boldsymbol{\Psi}_{SS}(k) \right]^{-1}, \quad (22)$$

where $\boldsymbol{\Psi}_{SS}(k)$ denotes the covariance of the near-end spectrum $\mathbf{S}(k)$ and A is a transition parameter [6]. $\boldsymbol{\Psi}_{b,\Delta\Delta}(k)$ above is the covariance of the temporal variations $\Delta \mathbf{W}_b(k)$ of the acoustic echo path. For this derivation of Kalman filter equations, we assumed that the echo path variation in different partitions is mutually uncorrelated and has zero mean, cf. [21].

3. PRACTICAL KALMAN FILTER IMPLEMENTATION

The above equations describe the theoretical framework for a partitioned-block implementation of the Kalman filter. Unfortunately, most of the $M \times M$ matrices involved are not diagonal. This makes an implementation in the AEC context rather unpractical. Moreover, the filter update in (18) does not assure the earlier imposed time-domain constraint on the adaptive filter coefficients for linear fast block convolution analogously to (7). This can lead to ambiguities and, thus, to potential convergence problems during the adaptation of the adaptive filter.

Analogously to [6], we therefore present a simplified version of the exact Kalman filter in this section. It turns out, that the simplified version can be used to derive a suitable step-size control for the update of each adaptive filter partition using the constrained version of the frequency domain least mean square (LMS) algorithm [14].

It is reasonable to assume that the covariance matrices $\boldsymbol{\Psi}_{ss}(k)$ and $\boldsymbol{\Psi}_{b,\Delta\Delta}(k)$ are diagonal. Analogously to [6] we further utilize approximations with respect to the excitation matrices $\mathbf{C}_b(k)$:

$$\mathbf{C}_b(k) \approx \frac{V}{M} \mathbf{X}_b(k), \quad (23)$$

$$\mathbf{C}_b(k) \mathbf{P}_b(k) \mathbf{C}_b^H(k) \approx \frac{V}{M} \mathbf{X}_b(k) \mathbf{P}_b(k) \mathbf{X}_b^H(k). \quad (24)$$

Substituting (24) and (24) into (21) and (22), we obtain a diagonalized version of the Kalman filter with

$$\mathbf{P}_b^+(k) = \left[\mathbf{I}_M - \frac{V}{M} \mathbf{K}_b(k) \mathbf{X}_b(k) \right] \mathbf{P}_b(k), \quad (25)$$

and the diagonalized Kalman gain

$$\mathbf{K}_b(k) = \mathbf{P}_b(k) \mathbf{X}_b^H(k) \left[\sum_{b=0}^{B-1} \mathbf{X}_b(k) \mathbf{P}_b(k) \mathbf{X}_b^H(k) + \frac{M}{V} \boldsymbol{\Psi}_{SS}(k) \right]^{-1} \quad (26)$$

respectively. The covariance matrix of the error spectrum $\mathbf{E}(k)$, given in (16), can be expressed as

$$\boldsymbol{\Psi}_{EE}(k) = \sum_{b=0}^{B-1} \mathbf{C}_b(k) \mathbf{P}_b(k) \mathbf{C}_b^H(k) + \boldsymbol{\Psi}_{SS}(k), \quad (27)$$

where we again assumed that the coefficients of different partitions are uncorrelated and have zero mean. If we substitute the covariance matrix of the error spectrum into (22), we obtain, together with (23), an alternative representation of the Kalman gain (26) according to

$$\mathbf{K}_b(k) = \frac{V}{M} \mathbf{P}_b(k) \mathbf{X}_b^H(k) \boldsymbol{\Psi}_{EE}^{-1}(k). \quad (28)$$

As shown in the following, we can immediately derive the partitioned-block frequency-domain adaptive filter in its constrained version [14] from the diagonalized version of the Kalman filter.

Noticing that the Kalman gain in its diagonalized form implicitly includes a step-size matrix, we express the Kalman gain according to

$$\mathbf{K}_b(k) = \boldsymbol{\mu}_b(k) \mathbf{X}_b^H(k), \quad (29)$$

where $\boldsymbol{\mu}_b(k)$ denotes the diagonal step-size matrix.

Regarding (18) and (20) and setting $A = 1$ for the update equation, the adaptation of the filter coefficients is then given by

$$\hat{\mathbf{W}}_b(k+1) = \hat{\mathbf{W}}_b(k) + \mathbf{G}_L \boldsymbol{\mu}_b(k) \mathbf{X}_b^H(k) \mathbf{E}(k), \quad (30)$$

where \mathbf{G}_L represents a pragmatic constraint matrix [14] to account for the time-domain zero-padding (constraining) in (7).

It is straightforward to verify from (28) and (29) that the desired step-size matrix $\boldsymbol{\mu}_b(k)$ is given by

$$\boldsymbol{\mu}_b(k) = \frac{V}{M} \mathbf{P}_b(k) \boldsymbol{\Psi}_{EE}^{-1}(k), \quad (31)$$

where it has been exploited that both, $\mathbf{X}_b(k)$ and $\boldsymbol{\Psi}_{EE}(k)$ are diagonal matrices.

The estimation of the system distance $\mathbf{P}_b(k)$ can be performed based on (19) and (21). Analogously to the single-partition case discussed in [6], we can determine the current coefficient error for each partition and frequency individually. Using (19), (25), and (29), we deduce a model-based recursive system distance estimator as

$$\begin{aligned} \mathbf{P}_b(k+1) &= A^2 \left(\mathbf{I}_M - \frac{V}{M} \boldsymbol{\mu}_b(k) \mathbf{X}_b^H(k) \mathbf{X}_b(k) \right) \mathbf{P}_b(k) \\ &\quad + \boldsymbol{\Psi}_{b,\Delta\Delta}(k). \end{aligned} \quad (32)$$

As proposed in [6], the innovation term $\boldsymbol{\Psi}_{b,\Delta\Delta}(k)$ is assumed to be a scaled version of the current covariance of the adaptive filter:

$$\boldsymbol{\Psi}_{b,\Delta\Delta}(k) = (1 - A^2) \boldsymbol{\Psi}_{b,\hat{W}}(m, k). \quad (33)$$

The matrix $\boldsymbol{\Psi}_{b,\hat{W}}(k)$ can be estimated from a temporally smoothed version of the squared magnitude of the adaptive filter coefficients,

$$\boldsymbol{\Psi}_{b,\hat{W}}(k) \approx \text{diag} \left\{ \hat{\mathbf{W}}_b(k) \right\} \text{diag} \left\{ \hat{\mathbf{W}}_b^H(k) \right\}. \quad (34)$$

Interestingly, the step-size matrix in (31) is very similar to the optimum step-size derived in [15] for minimizing the mean-square coefficient error. However, the proposed Kalman filter additionally provides the required system distance as an inherent building block.

4. RESIDUAL ECHO SUPPRESSION (RES) PART

In practice, the AEC is not always able to completely cancel the echo from the microphone signal. Even when the length of the adaptive filter can sufficiently capture the room impulse response, residual echoes remain due to the time-varying nature of the acoustic echo path and the presence of observation noise given by the near-end signal. In order to remove these residual echoes, a suppression filter $H(m, k)$ is commonly applied to the AEC output spectrum:

$$Z(m, k) = H(m, k) E(m, k). \quad (35)$$

It is well-known that the Wiener solution for the residual echo suppression filter $H(m, k)$ is given by, e.g., [3, 6],

$$H(m, k) = \frac{\Phi_{SS}(m, k)}{\Phi_{EE}(m, k)}, \quad (36)$$

where $\Phi_{SS}(m, k)$ and $\Phi_{EE}(m, k)$ denote the power spectral densities (PSD) of the near-end signal $s(n)$ and the AEC output $e(n)$.

Next, we present an interesting relation between the step-size parameters $\boldsymbol{\mu}_b(m, k)$ of the partitions, i.e., the elements on the diagonal of $\boldsymbol{\mu}_b(k)$, and the optimum residual echo suppression filter.

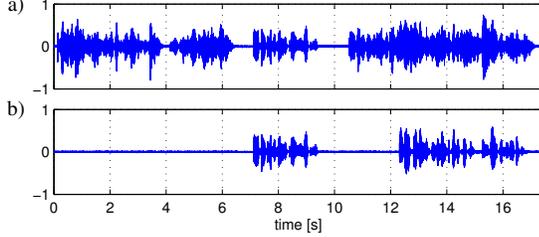


Fig. 2. a) Microphone signal, b) Near-end speech signal.

Obviously, the elements of the diagonal step-size matrix according to (31) can be written as

$$\mu_b(m, k) = \frac{V}{M} \frac{P_b(m, k)}{\Psi_{EE}(m, k)}. \quad (37)$$

As discussed in [6], the following relations between the PSDs and the elements on the main diagonal of the corresponding covariance matrices approximately hold:

$$\Psi_{SS}(m, k) = V\Phi_{SS}(m, k), \quad (38)$$

$$\Psi_{EE}(m, k) = V\Phi_{EE}(m, k), \quad (39)$$

$$P_b(m, k) = M\Phi_{b,r}(m, k). \quad (40)$$

Introducing (39) and (40) into (37) yields

$$\mu_b(m, k) = \frac{\Phi_{b,r}(m, k)}{\Phi_{EE}(m, k)}. \quad (41)$$

From (27), using the approximation (24), we further notice that the PSD of the error signal can be expressed in terms of the PSD of the coefficient error and the PSD of the near-end signal:

$$\Phi_{EE}(m, k) = \sum_{b=0}^{B-1} \Phi_{b,r}(m, k) |X_b(m, k)|^2 + \Phi_{SS}(m, k). \quad (42)$$

Considering (36), (41), and (42), we obtain a simple relation between the step-size parameters of each partition and the optimum echo suppression filter:

$$H(m, k) = 1 - \sum_{b=0}^{B-1} \mu_b(m, k) |X_b(m, k)|^2. \quad (43)$$

5. SIMULATION RESULTS

We present simulation results for the proposed echo control structure based on Kalman-AEC (K-AEC) and RES. The echo signal is simulated by convolving the loudspeaker signal (clean speech) with a measured room impulse response of an office with a reverberation time of about 300 ms. The signal-to-echo ratio during double-talk is about 0 dB and white noise has been added to the microphone to yield 25 dB near-end SNR. The sampling rate is 16 kHz. The AEC has been implemented with a partition-size equal to the frame shift, i.e., $L = R = 256$ samples. The DFT length is 512. The AEC uses $B = [1, 2, 5, 10, 15]$ partitions, which correspond to modeled time spans of [16, 32, 80, 160, 240] ms, respectively. Adaptive filters with up to 3840 taps are used here, which is significantly more than in previous Kalman filters [6, 20, 21], where less than 1000 taps were used. The transition parameter is $A = 0.999$.

The microphone and the near-end signal are illustrated in Figs. 2a and 2b, respectively. Fig. 3a depicts the system distances $\|\mathbf{w} - \hat{\mathbf{w}}(n)\|^2 / \|\mathbf{w}\|^2$ corresponding to the above simulation setup,

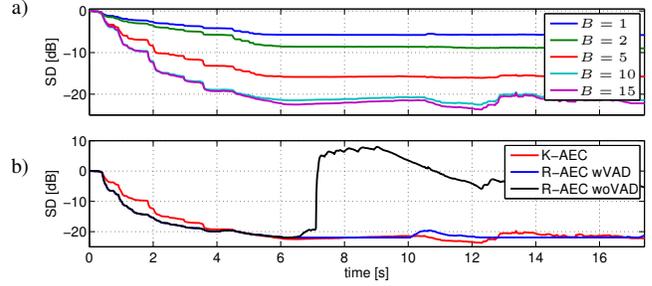


Fig. 3. System distances (SD) for a) K-AEC, with $B = [1, 2, 5, 10, 15]$ and b) K-AEC and R-AEC (with and without NE-VAD), with $B = 15$, obtained for an echo control scenario.

where $\hat{\mathbf{w}}(n)$ is the time-domain representation of the adaptive filter. As expected, the achievable system distance decreases with an increase in the number of partitions used.

Fig. 3b depicts the system distances for the K-AEC and a reference AEC (here referred to as R-AEC), where the stepsize is computed according to [23]. Two R-AEC variants are shown: 1) without a near-end voice activity detector (NE-VAD) and 2) with an ideal NE-VAD. The ideal NE-VAD avoids divergence of the R-AEC during double-talk or near-end single-talk by freezing the adaption when the near-end speaker is active. However, in practice the system distance of the R-AEC will be higher due to a non-ideal NE-VAD. The inherent step-size control in the K-AEC prevents the adaptive filter from divergence during double-talk or near-end single-talk, but it still allows the AEC to converge sufficiently fast to an accurate solution. Its performance is almost equivalent to the R-AEC with an ideal NE-VAD.

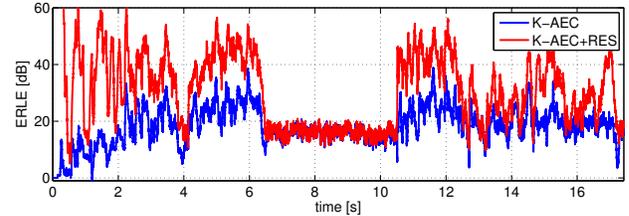


Fig. 4. ERLE at K-AEC and RES output, with $B = 15$.

Finally, Fig. 4 depicts the echo return loss enhancement (ERLE) [5] computed for the K-AEC and RES outputs. K-AEC provides an outstandingly stable baseline performance. The ERLE at the RES output is significantly higher than at the K-AEC output, especially during the far-end single-talk, which signifies the removal of the residual echo almost entirely. Although not shown here for brevity, the RES does not significantly affect the near-end signal.

6. CONCLUSIONS

The partitioned-block filter structure is often preferred in implementations of acoustic echo controllers to maintain simultaneous constraints on delay, computational complexity, memory requirements, and numerical stability. This paper adopted the larger framework of acoustic state-space modeling, which can comprehensively represent echo path variations and observation noise, to the partitioned-block structure. We found structural equivalence with multi-channel adaptive filter models known from stereophonic echo cancellation problems. We thus used the analogy to deduce, implement, and validate a complete echo cancellation and residual echo suppression system in state-space partitioned-block architecture.

7. REFERENCES

- [1] S. L. Gay and J. Benesty, Eds., *Acoustic Signal Processing for Telecommunications*, Kluwer Academic Publishers, 2000.
- [2] J. Benesty, T. Gänslér, D.R. Morgan, M.M. Sondhi, and S.L. Gay, *Advances in Network and Acoustic Echo Cancellation*, Springer, 2001.
- [3] G. Schmidt and E. Hänsler, *Acoustic echo and noise control: A practical approach*, Hoboken: Wiley, 2004.
- [4] S. Gustafsson, R. Martin, and P. Vary, "Combined acoustic echo control and noise reduction for hands-free telephony," *Signal Processing*, vol. 64, no. 1, pp. 21–32, January 1998.
- [5] G. Schmidt E. Hänsler, "Hands-free telephones – joint control of echo cancellation and postfiltering," *Signal Processing*, pp. 2295–2305, Nov. 2000.
- [6] G. Enzner and P. Vary, "Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones," *Signal Processing*, vol. 86, pp. 1140–1156, June 2006.
- [7] J. Wung, T.S. Wada, B.-H. Juang, B. Lee, T. Kalker, and R.W. Schafer, "A system approach to residual echo suppression in robust hands-free conferencing," in *Proc. of Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2011, pp. 445–448.
- [8] D.R. Morgan and J.C. Thi, "A delayless subband adaptive filter architecture," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 1819–1830, August 1995.
- [9] W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, New York, April 1988, pp. 2570–2573.
- [10] E.R. Ferrara, "Frequency-domain adaptive filtering," in *Adaptive Filters*, C.F.N. Cowan and P.M. Grant, Eds., pp. 145–179. Prentice Hall, Englewood Cliffs (NJ), 1985.
- [11] J. J. Shynk, "Frequency domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, pp. 14–37, Jan. 1992.
- [12] H. Buchner, J. Benesty, and W. Kellermann, "Generalized multichannel frequency-domain adaptive filtering: efficient realization and application to hands-free speech communication," *Signal Processing, Elsevier*, vol. 85, no. 3, pp. 549–570, 2005.
- [13] P. Vary and R. Martin, *Digital Speech Transmission - Enhancement, Coding, and Error Concealment*, John Wiley & Sons, Ltd., Chichester, England, 2006.
- [14] E. Moulines, O. A. Amrane, and Y. Grenier, "The generalized multidelay adaptive filter: Structure and convergence analysis," *IEEE Trans. on Signal Processing*, vol. 43, no. 1, pp. 14–28, Jan. 1995.
- [15] B. H. Nitsch, "A frequency-selective stepfactor control for an adaptive filter algorithm working in the frequency domain," *Signal Processing*, vol. 80, no. 9, pp. 1733–1745, Sept. 2000.
- [16] P.C.W. Sommen, *Adaptive Filtering Methods: On methods to use a priori information in order to reduce complexity while maintaining convergence properties*. Ph.D. thesis, Technical University of Eindhoven, ISBN 90-9005143-0, 1992.
- [17] J.-S. Soo and K.K. Pang, "Multidelay block frequency domain adaptive filter," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 373–376, February 1990.
- [18] G. Enzner and P. Vary, "A soft-partitioned frequency-domain adaptive filter for acoustic echo cancellation," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Hong Kong (China), May 2003, pp. 393–396.
- [19] M.-A. Jung and T. Fingscheidt, "A shadow filter approach to a wideband FDAF-based automotive handsfree system," in *Proc. of 5th Biennial Workshop on DSP for In-Vehicle Systems*, Kiel, Sept. 2011.
- [20] S. Malik and G. Enzner, "Recursive Bayesian control of multichannel acoustic echo cancellation," *IEEE Signal Processing Letters*, vol. 18, no. 11, pp. 619–622, Nov. 2011.
- [21] S. Malik and G. Enzner, "State-space frequency-domain adaptive filtering for nonlinear acoustic echo cancellation," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 20, no. 7, pp. 2065–2079, Sept. 2012.
- [22] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms and Applications*, New Jersey: Prentice Hall, 1996.
- [23] S. Haykin, *Adaptive Filter Theory*, New Jersey: Prentice Hall, 1996.