MODEL-BASED VS. TRADITIONAL FREQUENCY-DOMAIN ADAPTIVE FILTERING IN THE PRESENCE OF CONTINUOUS DOUBLE-TALK AND ACOUSTIC ECHO PATH VARIABILITY

Sarmad Malik and Gerald Enzner

Institute of Communication Acoustics (IKA)
Ruhr-University Bochum, D-44780 Bochum, Germany
E-mail:{sarmad.malik|gerald.enzner}@ruhr-uni-bochum.de

ABSTRACT

This contribution presents the utilization of a linear dynamical system as an echo path model for the acoustic echo cancellation (AEC) problem. This comprehensive model incorporates variations in the echo path, near-end speech, and the ambient noise. It is well known that the optimum state estimation of such a dynamical system can be performed by the Kalman filtering algorithm. We show that the frequency-domain implementation of the Kalman filter resembles an LMS-type frequency-domain adaptive filter (FDAF), with the Kalman gain as an elegant substitute for the otherwise heuristic step-size. Owing to the model-based approach, the computation of the Kalman gain inherently includes the near-end signal statistics as well as the echo path variability. This renders the adaptive echo cancellation process fast and robust in the presence of double talk and noise.

***Index Terms*—** Echo cancellation, Frequency-domain adaptive filter, State-space modeling, Kalman filter

1. INTRODUCTION

Algorithms designed for AEC have to consider several design aspects such as robustness in the presence of near-end signal, convergence speed, tracking ability and low complexity. Robustness in the presence of double-talk requires an effective step-size control, which necessitates the inclusion of the near-end signal statistics in the adaptation process. Numerous variable step-size algorithms have been proposed in the past and just recently, e.g., [1, 2] (and references therein). The estimation of the near-end signal statistics has turned out to be problematic, therefore double-talk detectors (DTDs) are frequently used to complement the schemes [2]. Nevertheless, it is important to realize that on the onset of double-talk, the adaptive algorithm diverges significantly before the double-talk is detected. Due to this detection lag and other inherent DTD imperfections, a DTD-free solution must be sought.

It is essential to emphasize here that the variable step-size control and the inclusion of the near-end statistics should not be an outcome of a heuristic deliberation. Any suboptimal solution, e.g., using excessive regularization, classification or detection, may cause an increased misadjustment and slower convergence than desired [3, 4].

General advantages of frequency-domain implementations of adaptive algorithms over time-domain implementations have been duly mentioned in the literature [5]. The objective of this paper is to direct the attention to a Kalman filtering solution for AEC in DFT domain, where the problem is viewed considering a state-space model for the echo path [6]. The model-based optimal system identification inherently includes variable step-size control and the near-end signal statistics, and thus omits any need for a DTD. By means of appropriate approximations in DFT domain, a diagonalized set of frequency-domain Kalman update equations can be derived. In Section 2, we will show that the diagonalized Kalman filter acquires a structural form for which we introduce the term "State-Space FDAF". Structural similarities and differences between the classical FDAF and the State-Space FDAF will be clarified, which will help to explain the inherent robustness of State-Space FDAF. The superiority of the model-based algorithm will be further confirmed by simulation results, in Section 3, in the presence of continuous double-talk and time-varying echo path.

Throughout the paper we use nonbold lowercase letters for time-domain scalar quantities, bold lowercase for vectors and bold uppercase for frequency-domain quantities. The superscript H denotes Hermitian transposition. We use $\mathbf{F}_M$ to denote the DFT matrix of size M. Lowercase letter "k" and "κ" are reserved for time and block-time indices, respectively.

2. SIGNAL MODELS AND ALGORITHMS

A general AEC scenario occurring in a loudspeaker-enclosure-microphone (LEM) system is shown in Fig. 1. The far-end speech signal $x(k)$ is transmitted through the linear echo path $\mathbf{w}(k)$ and gives the echo signal $d(k)$. The received signal at the microphone, $y(k)$, is the sum of the near-end signal $s(k)$ and the echo signal $d(k)$. The near-end signal comprises the near-end speech signal and ambient noise. The estimated echo path by the adaptive algorithm is denoted by $\widehat{\mathbf{w}}(k)$.

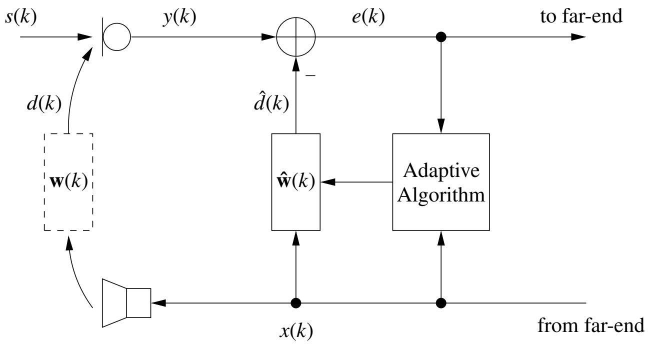

Fig. 1. Acoustic front-end, e.g., of a hands-free telephone.

2.1. Overlap-Save Echo Path Model

A formulation of the linear convolution echo path model by means of the overlap-save method requires block-oriented definitions of signals. We define the excitation vector $\mathbf{x}(\kappa)$, observation vector $\mathbf{y}(\kappa)$, and observation noise vector $\mathbf{s}(\kappa)$:

$$\begin{aligned} \mathbf{x}(\kappa) &= [x(\kappa R - M + 1), x(\kappa R - M + 2), ..., x(\kappa R)]^H \\ \mathbf{y}(\kappa) &= [y(\kappa R - R + 1), y(\kappa R - R + 2), ..., y(\kappa R)]^H \\ \mathbf{s}(\kappa) &= [s(\kappa R - R + 1), s(\kappa R - R + 2), ..., s(\kappa R)]^H \end{aligned}, \quad (1)$$

where R denotes the frame-shift. A complex-valued $M \times M$ excitation matrix $\mathbf{X}(\kappa)$ is created by first applying DFT and then diagonalization to the far-end signal:

$$\mathbf{X}(\kappa) = diag\left\{\mathbf{F}_M \mathbf{x}(\kappa)\right\}. \quad (2)$$

A frequency-domain echo path vector $\mathbf{W}(\kappa)$ is obtained on the basis of the overlap-save constraint that only $M - R$ non-zero coefficients of the time-domain echo path $\mathbf{w}(\kappa)$ can be represented,

$$\mathbf{W}(\kappa) = \mathbf{F}_M \begin{pmatrix} \mathbf{w}(\kappa) \\ \mathbf{0} \end{pmatrix}, \quad (3)$$

where

$$\mathbf{w}(\kappa) = \left[w_0(\kappa), w_1(\kappa), ..., w_{M-R-1}(\kappa)\right]^H. \quad (4)$$

Based on these definitions, we can express the linear additive echo path model in Fig. 1 in a compact matrix-vector form entirely in the frequency-domain:

$$\begin{aligned} \mathbf{Y}(\kappa) &= \mathbf{F}_M \mathbf{Q} \mathbf{y}(\kappa) \\ &= \mathbf{F}_M \mathbf{Q} \mathbf{s}(\kappa) + \mathbf{F}_M \mathbf{Q} \mathbf{Q}^H \mathbf{F}_M^{-1} \mathbf{X}(\kappa) \mathbf{W}(\kappa) \\ &= \mathbf{S}(\kappa) + \mathbf{C}(\kappa) \mathbf{W}(\kappa), \end{aligned} \quad (5)$$

where $\mathbf{Q}^H = (\mathbf{0}\ \mathbf{I}_R)$ is an $R \times M$ projection matrix, included to linearize the cyclic convolution in DFT domain, and $\mathbf{I}_R$ denotes an $R \times R$ identity matrix. The term $\mathbf{C}(\kappa)\mathbf{W}(\kappa)$ in (5) is the echo signal vector in the frequency-domain, where $\mathbf{C}(\kappa) = \mathbf{F}_M \mathbf{Q} \mathbf{Q}^H \mathbf{F}_M^{-1} \mathbf{X}(\kappa)$. Here, we can further express that $\mathbf{C}(\kappa) = \mathbf{G}\mathbf{X}(\kappa)$ such that $\mathbf{G} = \mathbf{F}_M \mathbf{Q} \mathbf{Q}^H \mathbf{F}_M^{-1}$ is constant.

2.2. Frequency-Domain Adaptive Filter

Frequency-domain adaptive filtering is a traditional solution for system identification in AEC [5, 7]. Using the introduced notation, the error signal computation and the update equation for the estimated echo path $\widehat{\mathbf{W}}(\kappa)$ are given as:

$$\begin{aligned} \mathbf{E}(\kappa) &= \mathbf{F}_M \mathbf{Q}[\mathbf{y}(\kappa) - \mathbf{Q}^H \mathbf{F}_M^{-1} \mathbf{X}(\kappa) \widehat{\mathbf{W}}(\kappa)] \\ \widehat{\mathbf{W}}(\kappa + 1) &= \widehat{\mathbf{W}}(\kappa) + \boldsymbol{\mu}(\kappa) \mathbf{X}^H(\kappa) \mathbf{E}(\kappa). \end{aligned} \quad (6)$$

In (6), $\boldsymbol{\mu}(\kappa)$ is an $M \times M$ diagonal matrix with each entry representing an individual step-size parameter for each frequency bin. The traditional step-size parameter matrix is given as:

$$\boldsymbol{\mu}(\kappa) = \alpha \boldsymbol{\Psi}_{XX}^{-1}(\kappa), \quad (7)$$

where $\boldsymbol{\Psi}_{XX}(\kappa)$ and α are the power spectral density estimate of the far-end signal and the adaptation constant in the range $0 < \alpha < 1$, respectively. The $M \times M$ diagonal matrix $\boldsymbol{\Psi}_{XX}(\kappa)$ is computed by recursive averaging,

$$\boldsymbol{\Psi}_{XX}(\kappa) = \gamma \boldsymbol{\Psi}_{XX}(\kappa - 1) + (1 - \gamma) \mathbf{X}^H(\kappa) \mathbf{X}(\kappa), \quad (8)$$

where γ is a forgetting factor in the range $0 < \gamma < 1$. Moreover, the cyclic correlation operation in the update equation for the estimated echo path $\widehat{\mathbf{W}}(\kappa)$ in (6) can be linearized by applying a *gradient constraint* [7].

2.3. Markov Model of the Time-Varying Echo Path

A pragmatic approach towards acoustic echo control entails that the variation in the echo path between time instants κ and $\kappa + 1$ be considered smooth and gradual. Therefore, a first-order Markov model [7] can be employed to express the time-varying behavior of the echo path $\mathbf{W}(\kappa)$ [6]:

$$\mathbf{W}(\kappa + 1) = A \cdot \mathbf{W}(\kappa) + \Delta \mathbf{W}(\kappa). \quad (9)$$

In (9), A denotes a time-invariant state transition coefficient and $\Delta\mathbf{W}(\kappa)$ represents a zero-mean, independent and uncorrelated process noise vector, accommodating for the uncertainty in the echo path variation. It is assumed that the value of A is less than but close to unity. Combining the frequency-domain observation model (5), and the first order Markov model (9), we obtain the comprehensive state-space model as represented by the signal flow graph in Fig. 2.

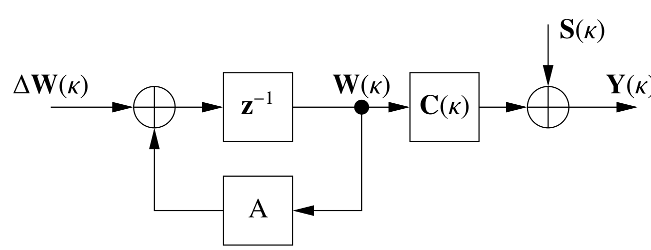

Fig. 2. Stochastic state-space model of the unknown time-varying echo path $\mathbf{W}(\kappa)$ in the block frequency-domain.

2.4. Exact Kalman Filter in DFT Domain

The one-step prediction based Kalman filtering solution for obtaining the estimate $\widehat{\mathbf{W}}(\kappa)$ of the echo path $\mathbf{W}(\kappa)$, subject to the state-space model in Fig. 2, can be summarized as [4]:

$$\begin{aligned}
\mathbf{K}(\kappa) &= A\cdot\mathbf{P}(\kappa-1)\mathbf{C}^H(\kappa)[\mathbf{C}(\kappa)\mathbf{P}(\kappa-1)\mathbf{C}^H(\kappa)+\mathbf{\Psi}_{SS}(\kappa)]^{-1}\\
\mathbf{E}(\kappa) &= \mathbf{F}_M\mathbf{Q}[\mathbf{y}(\kappa)-\mathbf{Q}^H\mathbf{F}_M^{-1}\mathbf{X}(\kappa)\widehat{\mathbf{W}}(\kappa-1)]\\
\widehat{\mathbf{W}}(\kappa) &= A\cdot\widehat{\mathbf{W}}(\kappa-1)+\mathbf{K}(\kappa)\mathbf{E}(\kappa)\\
\mathbf{P}(\kappa) &= A\cdot[A\cdot\mathbf{I}_M-\mathbf{K}(\kappa)\mathbf{C}(\kappa)]\mathbf{P}(\kappa-1)+\mathbf{\Psi}_{\Delta\Delta},
\end{aligned} \tag{10}$$

where $\mathbf{K}(\kappa)$ is the Kalman gain, $\mathbf{E}(\kappa)$ is the error signal or innovation process, $\widehat{\mathbf{W}}(\kappa)$ is the estimate of the echo path state, $\mathbf{P}(\kappa)$ is the state estimation error covariance, $\mathbf{I}_M$ is an $M \times M$ identity, $\mathbf{\Psi}_{\Delta\Delta} = E[\Delta\mathbf{W}(\kappa)\Delta\mathbf{W}^H(\kappa)]$ is the process noise covariance matrix, and $\mathbf{\Psi}_{SS}(\kappa) = E[\mathbf{S}(\kappa)\mathbf{S}^H(\kappa)]$ is the near-end signal covariance matrix with $E[\ \cdot\ ]$ denoting the expectation operation.

In the Kalman filter, it is crucial to observe that $\mathbf{P}(\kappa)$, evaluated from the *Riccati difference equation*, acts as a system distance between estimated and true echo path and essentially controls the value of the Kalman gain $\mathbf{K}(\kappa)$.

2.5. State-Space FDAF

It has been verified that the term $\mathbf{G}$, as introduced in Section 2.1, can be approximated as a scaled identity, $\mathbf{G} \approx \frac{R}{M}\mathbf{I}_M$, and therefore $\mathbf{C}(\kappa) \approx \frac{R}{M}\mathbf{X}(\kappa)$, while $\mathbf{C}(\kappa)\mathbf{P}(\kappa-1)\mathbf{C}^H(\kappa) \approx \frac{R}{M}\mathbf{X}(\kappa)\mathbf{P}(\kappa-1)\mathbf{X}^H(\kappa)$ [6, 8]. Furthermore, process noise and near-end signal covariance matrices maintain nearly diagonal attributes in the transform domain. If the state estimation error covariance matrix is initialized as a diagonal matrix $\mathbf{P}(0)$, then the terms $\mathbf{K}(\kappa)$ and $\mathbf{P}(\kappa)$ are automatically diagonalized, too. This enables a fast frequency-domain implementation of the Kalman filter using vector arithmetics and FFT/IFFT. Diagonalization further renders the matrix products commutative, whereby the Kalman gain in (10) can be expressed as

$$\mathbf{K}(\kappa)=A\cdot\mathbf{P}(\kappa-1)[\mathbf{X}(\kappa)\mathbf{P}(\kappa-1)\mathbf{X}^H(\kappa)+\frac{M}{R}\mathbf{\Psi}_{SS}(\kappa)]^{-1}\mathbf{X}^H(\kappa) \tag{11}$$

and thus we may rewrite the echo path update equation as

$$\widehat{\mathbf{W}}(\kappa)=A\cdot\widehat{\mathbf{W}}(\kappa-1) + \boldsymbol{\mu}_K(\kappa)\mathbf{X}^H(\kappa)\mathbf{E}(\kappa), \tag{12}$$

where $\boldsymbol{\mu}_K(\kappa)$ is the optimum Kalman step-size given by:

$$\boldsymbol{\mu}_K(\kappa)=A\cdot\mathbf{P}(\kappa-1)[\mathbf{X}(\kappa)\mathbf{P}(\kappa-1)\mathbf{X}^H(\kappa)+\frac{M}{R}\mathbf{\Psi}_{SS}]^{-1}. \tag{13}$$

From (12), a cross-correlation operation is evident and in analogy to the FDAF [7], an additional gradient constraint could be applied to linearize the cyclic correlation in the DFT domain. Using (12) and (13), we formally obtain a state-space frequency-domain adaptive structure manifesting our overlap-save echo path model, as shown in Fig. 3. The comparison between (7) and (13) highlights that the computation of the Kalman step-size involves not only the far-end statistics, but also the state estimation error and near-end signal covariances. An approximation of the near-end statistics $\mathbf{\Psi}_{SS}(\kappa)$ can be retrieved from the error signal $\mathbf{E}(\kappa)$ [4].

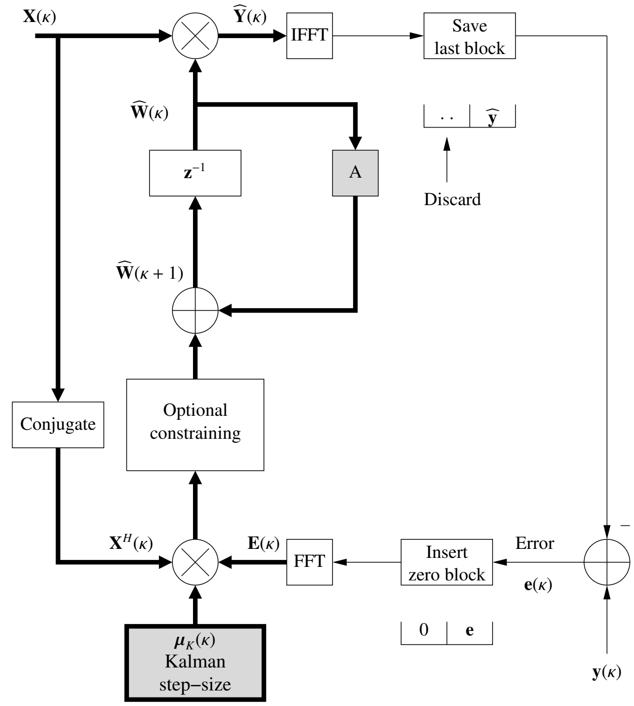

Fig. 3. State-Space FDAF.

3. RESULTS AND COMPARISON

We will present simulation results pertaining to fixed and time-varying echo paths. All comparisons will be carried out between the State-Space FDAF and the traditional FDAF. The performance especially in the simultaneous presence of double-talk and echo path variability will be highlighted. Furthermore, as the State-Space FDAF performs system identification based on the aforementioned Markov model, model mismatch will also be investigated.

We consider echo path impulse responses comprising of 512 taps. The adaptive filter length is set to $M-R = 512$, too, with a block size of $M = 1024$. The near-end-to-echo signal power ratio in double-talk is set to 0 dB. The signal presented in Fig. 4(a) is used as the near-end speech signal. Owing to the absence of any specific assumption in our algorithm regarding the autocorrelation and stationarity properties of the far-end signal, we use a stationary white Gaussian signal in the following. The signal presented in Fig. 4(b) is the echo signal obtained for a time-varying echo path simulation. In all scenarios, single-talk and double-talk, the stationary near-end ambient noise level is set to about -30 dB. The criterion we use for measuring performance and carrying out compar-

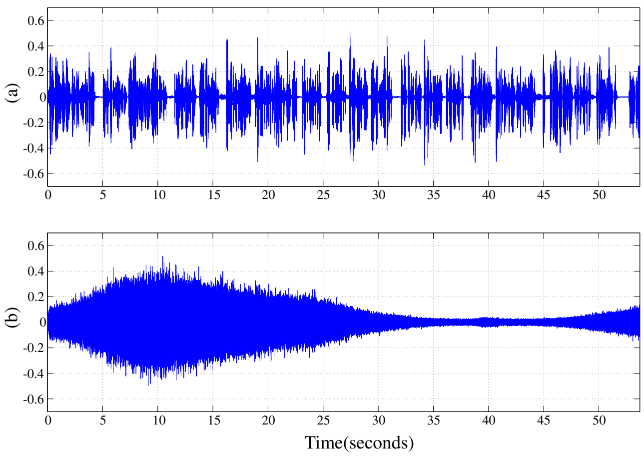

Fig. 4. (a) Near-end speech; (b) Echo for a time-varying echo path.

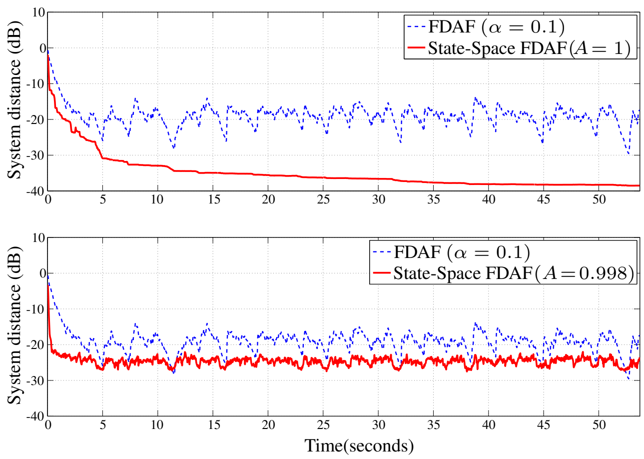

Fig. 5. Fixed echo path simulation with continuous double-talk.

isons of adaptive algorithms is the *relative system distance*, $10\log_{10}(||\mathbf{W}(\kappa) - \widehat{\mathbf{W}}(\kappa)||^2/||\mathbf{W}(\kappa)||^2)$.

Simulation results for the fixed echo path in the presence of continuous double-talk are presented in Fig. 5. It is evident that State-Space FDAF clearly outperforms FDAF in terms of convergence rate and final misadjustment for $A = 1$. The setting $A = 1$, corresponds to a fixed echo path in the state-space model (9). For $A = 0.998$, the performance of State-Space FDAF approaches closer to FDAF, which is due to the fact that the echo path identification is model-based and the model mismatch has an impact on the results.

In Fig. 6, we analyze the behavior for a time-varying echo path. In Fig. 6 (Upper), only a comparison of the tracking ability is sought. We thus consider a far-end single-talk case and clearly observe that State-Space FDAF shows a lower system distance as compared to the traditional FDAF. In Fig. 6 (Lower), performances are compared in the presence of continuous double-talk, which constitutes the harshest experiment. Naturally, State-Space FDAF exhibits a few dB of principle loss as compared to the single-talk case, but its performance is still remarkably and consistently superior to FDAF.

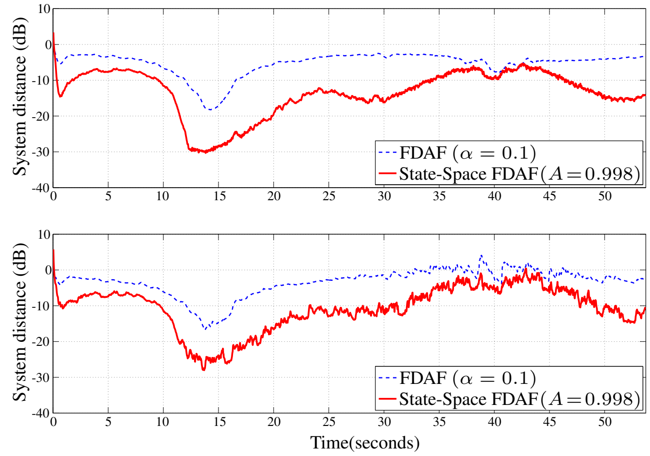

Fig. 6. Time-varying echo path simulation: (Upper) Single-talk; (Lower) Double-talk.

4. CONCLUSIONS

A state-space frequency-domain adaptive filtering approach has been discussed and compared to the traditional version in terms of the resulting system distance. Since the underlying state-space model of the echo path features the existence of double-talk and echo path variability, State-Space FDAF provides all structural elements needed for fast adaptation and inherent robustness. Nevertheless, the algorithm is efficient and easy to implement. Simulations have consistently demonstrated the practical benefit of the discussed algorithm and the underlying model in the presence of double-talk, ambient noise, and continuously changing echo paths.

5. REFERENCES

[1] M. A. Iqbal and S. L. Grant, "Novel variable step size NLMS algorithms for echo cancellation," *Proc. ICASSP*, pp. 241–244, Mar. 2008.

[2] C. Paleologu, S. Ciochina, and J. Benesty, "Double-talk robust VSS-NLMS algorithm for under-modeling acoustic echo cancellation," *Proc. ICASSP*, pp. 245–248, Mar. 2008.

[3] T. van Watershoot, G. Rombouts, P. Verheove, and M. Moonen, "Double-talk robust prediction error identification algorithms for acoustic echo cancellation," *IEEE Trans. Signal Processing*, vol. 55, no. 3, pp. 846–858, Mar. 2007.

[4] G. Enzner, *A Model-Based Optimum Filtering Approach to Acoustic Echo Control: Theory and Practice*, Wissenschaftsverlag Mainz, Aachen, Apr. 2006.

[5] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992.

[6] G. Enzner and P. Vary, "Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones," *Signal Processing, Elsevier*, vol. 86, no. 6, pp. 1140–1156, Jun. 2006.

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

[8] J. Benesty, T. Gaensler, D. Morgan, M. Sondhi, and S. Gay, *Advances in Network and Acoustic Echo Cancellation*, Springer, Berlin, 2001.