VARIATIONAL BAYESIAN INFERENCE FOR NONLINEAR ACOUSTIC ECHO CANCELLATION USING ADAPTIVE CASCADE MODELING

Sarmad Malik and Gerald Enzner

Institute of Communication Acoustics, Ruhr-Universität Bochum, Germany
{sarmad.malik,gerald.enzner}@rub.de

ABSTRACT

In this contribution, we present a variational Bayesian framework for the acoustic echo cancellation problem in the presence of a memoryless loudspeaker nonlinearity. We pursue a cascade modeling strategy, where first-order Markov models are described over the acoustic echo path and the nonlinear expansion coefficients. An iterative algorithm is then derived that learns the posterior on the echo path and the nonlinear coefficients to fit the evidence distribution. We show that the formulated variational Bayesian state-space frequency-domain adaptive filter is efficiently implementable and performs joint learning of the echo path and the loudspeaker nonlinearity. The algorithm exploits the internal exchange of the reliability information, resulting in effective linear and nonlinear echo cancellation.

Index Terms— Adaptive filtering, cascade modeling, frequency-domain, nonlinear echo cancellation, state-space, variational Bayes.

1. INTRODUCTION

Over the years, the problem of nonlinear echo cancellation has been addressed by means of equivalent multichannel [1, 2] and cascade structures [3, 4] based upon memoryless Hammerstein modeling. In this context, multichannel power filters with normalized least-mean-square (NLMS)-type adaptation have been described that incorporate a necessary adaptive orthogonalization mechanism [1]. A cascade solution with recursive least-squares (RLS) estimation of a memoryless polynomial pre-processor has been proposed in [3]. Here, the RLS estimation is carried out in conjunction with the NLMS-type adaptive learning of the acoustic echo path.

In our work, we resolve the nonlinearity in the Hammerstein model by a basis-generic expansion [2] that is weighted by corresponding nonlinear expansion coefficients. It is essential to realize that the simultaneous learning of the nonlinear expansion coefficients and the linear FIR echo path in the Hammerstein model constitutes a joint estimation scenario. In order to derive a robust and recursive algorithm, we adopt a novel approach and model the two unknown quantities as independent random variables with a first-order Markov property. This inherently enables the resulting Kalman-type algorithm to compute the uncertainty measures of the quantities of interest, and incorporate them in the adaptation process. The Markov modeling of the two variables of interest followed by the inclusion of the near-end observation noise into the estimation framework, yields a *composite* state-space model for nonlinear echo cancellation.

We revert to a variational Bayesian methodology [5] for inferring the composite state-space model, which we have formulated in the DFT-domain. The ensuing derivation results in our variational Bayesian state-space frequency-domain adaptive filter (VB-SSFDAF) that learns the posterior distributions of the unknown quantities to fit the *evidence* distribution. The joint learning framework facilitates the exchange of reliability measures among the partial posterior estimators that are derived for the nonlinear coefficients and the echo path, respectively. We demonstrate the VB-SSFDAF to be a stable and efficiently implementable recursive algorithm for nonlinear echo cancellation.

In Sec. 2, we describe the DFT-domain composite state-space model. The derivation of the variational algorithm, i.e., VB-SSFDAF, is presented in Sec. 3. In Sec. 4, we support the derived algorithm with simulation results for single and double-talk cases. We finally conclude our work in Sec. 5.

2. SYSTEM MODEL

In Fig. 1, which depicts a nonlinear echo cancellation scenario, the input signal x_t from the far-end undergoes a nonlinear transformation due to a loudspeaker nonlinearity $f(\cdot)$ to give the nonlinearly mapped input signal $f(x_t)$. Here, t denotes the sample-time index. The nonlinearly mapped input signal $f(x_t)$ then gets linearly convolved with the echo path $\mathbf{w}'_t$ to generate the echo signal d_t. The echo signal d_t is superimposed with the near-end speech and noise s_t resulting in the microphone signal y_t. The aim then of the adaptive algorithm is to come up with estimates of the echo path $\widehat{\mathbf{w}}'_t$ and the nonlinear mapping $\widehat{f}(\cdot)$ such that the inferred echo signal $\widehat{d}_t$ adequately cancels the actual echo signal d_t.

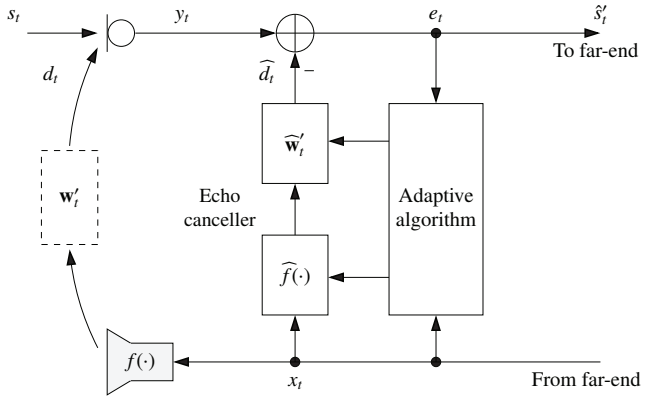

Fig. 1. Acoustic echo cancellation in the presence of a loudspeaker nonlinearity $f(\cdot)$.

This work was carried out in cooperation with Nokia.

2.1. DFT-Domain Observation Model

We initiate our derivation by resolving the nonlinearly mapped input signal $f(x_t)$ in terms of a basis-generic expansion

$$f(x_t) = \sum_{i=1}^{p} a'_{t,i}\phi_i(x_t) \ , \tag{1}$$

where $\phi_i(\cdot)$ is the ith basis function and p is the considered expansion order. As we aim to derive an adaptive filtering algorithm in the DFT-domain, it is necessary to describe the frame-based definition of the input signal,

$$\mathbf{x}_\tau = [\ x_{\tau R-M+1}\ x_{\tau R-M+2}\ \ldots\ x_{\tau R}\ ]^{\mathrm{T}} \ , \tag{2}$$

where R and M are the frame-shift and the frame-size, respectively. The symbol T denotes transposition. Using (1) and (2), we express the frame-based nonlinearly mapped input signal $\mathbf{f}_\tau$ as

$$\mathbf{f}_\tau = [\ f(x_{\tau R-M+1})\ f(x_{\tau R-M+2})\ \ldots\ f(x_{\tau R})\ ]^{\mathrm{T}} \ , \tag{3}$$

and substitute (1) into (3) to get

$$\begin{aligned}\mathbf{f}_\tau &= \left[\ \sum_{i=1}^{p} a_{\tau,i}\phi_i(x_{\tau R-M+1})\ \ldots\ \sum_{i=1}^{p} a_{\tau,i}\phi_i(x_{\tau R})\ \right]^{\mathrm{T}} \\ &= \sum_{i=1}^{p} a_{\tau,i}\mathbf{x}_{\tau,i} \ ,\end{aligned} \tag{4}$$

where $\mathbf{x}_{\tau,i} = [\phi_i(x_{\tau R-M+1})\ \phi_i(x_{\tau R-M+2})\ \ldots\ \phi_i(x_{\tau R})]^{\mathrm{T}}$. The frame-time index is denoted by τ, and $a_{\tau,i} = a'_{t=\tau R,i}$ is the frame-based version of the nonlinear coefficients $a'_{t,i}$. The term $\mathbf{f}_\tau$ in (4) is converted into the $M \times M$ DFT-domain matrix $\widetilde{\mathbf{X}}_\tau$ by applying an $M \times M$ DFT-matrix $\mathbf{F}_M$ followed by diagonalization

$$\widetilde{\mathbf{X}}_\tau = \mathrm{diag}\,\{\mathbf{F}_M\mathbf{f}_\tau\} = \sum_{i=1}^{p} a_{\tau,i}\mathbf{X}_{\tau,i} \ . \tag{5}$$

Here, $\mathbf{X}_{\tau,i} = \mathrm{diag}\,\{\mathbf{F}_M\mathbf{x}_{\tau,i}\}$ is the ith component of the nonlinearly mapped input signal in the DFT-domain.

We then consider $L = M - R$ non-zero coefficients of the frame-based echo path $\mathbf{w}_\tau = \mathbf{w}'_{t=\tau R}$ and apply the DFT, i.e.,

$$\mathbf{W}_\tau = \mathbf{F}_M \begin{bmatrix}\mathbf{w}_\tau^{\mathrm{T}} & \mathbf{0}_{R\times 1}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \ , \tag{6}$$

to obtain the $M \times 1$ frequency-domain vector $\mathbf{W}_\tau$, where $\mathbf{0}_{R\times 1}$ denotes a zero-padding operation. Using (5) and (6), the frame-based microphone signal $\mathbf{y}_\tau = [y_{\tau R-R+1}\ y_{\tau R-R+2}\ \ldots\ y_{\tau R}]^{\mathrm{T}}$ can be expressed via the overlap-save convolution,

$$\mathbf{y}_\tau = \boldsymbol{\Upsilon}^{\mathrm{T}}\mathbf{F}_M^{-1}\widetilde{\mathbf{X}}_\tau\mathbf{W}_\tau + \mathbf{s}_\tau \ . \tag{7}$$

The matrix $\boldsymbol{\Upsilon}^{\mathrm{T}} = [\mathbf{0}_{R\times L}\ \ \mathbf{I}_R]$ is an $R \times M$ projection matrix and the near-end disturbance vector $\mathbf{s}_\tau$ is defined analogous to $\mathbf{y}_\tau$. We zero-padd $\mathbf{y}_\tau$ using $\boldsymbol{\Upsilon}$ and apply the Fourier matrix $\mathbf{F}_M$, i.e., $\mathbf{Y}_\tau = \mathbf{F}_M\boldsymbol{\Upsilon}\mathbf{y}_\tau$, to get the DFT-domain observation model,

$$\begin{aligned}\mathbf{Y}_\tau &= \mathbf{F}_M\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^{\mathrm{T}}\mathbf{F}_M^{-1}\widetilde{\mathbf{X}}_\tau\mathbf{W}_\tau + \mathbf{F}_M\boldsymbol{\Upsilon}\mathbf{s}_\tau && (8)\\ &= \mathbf{G}\widetilde{\mathbf{X}}_\tau\mathbf{W}_\tau + \mathbf{S}_\tau && (9)\\ &= \widetilde{\mathbf{C}}_\tau\mathbf{W}_\tau + \mathbf{S}_\tau \ , && (10)\end{aligned}$$

where $\widetilde{\mathbf{C}}_\tau = \mathbf{G}\widetilde{\mathbf{X}}_\tau$ and $\mathbf{G} = \mathbf{F}_M\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^{\mathrm{T}}\mathbf{F}_M^{-1}$. The term $\mathbf{S}_\tau = \mathbf{F}_M\boldsymbol{\Upsilon}\mathbf{s}_\tau$ is modeled as zero-mean and bin-wise uncorrelated Gaussian noise with an $M \times M$ diagonal covariance matrix $\boldsymbol{\Psi}_\tau^S = \langle\mathbf{S}_\tau\mathbf{S}_\tau^{\mathrm{H}}\rangle$. The symbol H denotes Hermitian transposition and $\langle\cdot\rangle$ is the expectation operator.

2.2. Composite State-Space Model

It can be deduced from (5) and (10) that $\widetilde{\mathbf{C}}_\tau = \sum_{i=1}^{p} a_{\tau,i}\mathbf{C}_{\tau,i}$, where $\mathbf{C}_{\tau,i} = \mathbf{G}\mathbf{X}_{\tau,i}$ is the overlap-save contrained version of the ith component $\mathbf{X}_{\tau,i}$ of the nonlinearly mapped input signal $\widetilde{\mathbf{X}}_\tau$ in the DFT-domain. We substitute $\widetilde{\mathbf{C}}_\tau = \sum_{i=1}^{p} a_{\tau,i}\mathbf{C}_{\tau,i}$ into (10) to get

$$\begin{aligned}\mathbf{Y}_\tau &= \sum_{i=1}^{p} a_{\tau,i}\mathbf{C}_{\tau,i}\mathbf{W}_\tau + \mathbf{S}_\tau \\ &= [\ \mathbf{C}_{\tau,1}\ \ldots\ \mathbf{C}_{\tau,p}\ ]\,[\ a_{\tau,1}\mathbf{I}_M \ldots a_{\tau,p}\mathbf{I}_M\ ]^{\mathrm{T}}\,\mathbf{W}_\tau + \mathbf{S}_\tau \\ &= \underline{\mathbf{C}}_\tau\,[\ \mathbf{a}_\tau \otimes \mathbf{I}_M\ ]\,\mathbf{W}_\tau + \mathbf{S}_\tau \ ,\end{aligned} \tag{11}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{a}_\tau = [a_{\tau,1}\ \ldots\ a_{\tau,p}]^{\mathrm{T}}$. In (11), $\underline{\mathbf{C}}_\tau = [\mathbf{C}_{\tau,1}\ldots\mathbf{C}_{\tau,p}]$. The terms $\mathbf{W}_\tau$ and $\mathbf{a}_\tau$ represent the unknown echo path and the nonlinear expansion coefficients, respectively. Here, we introduce another useful form of the DFT-domain observation that is fully equivalent to (10) and (11)

$$\mathbf{Y}_\tau = \underline{\mathbf{C}}_\tau\,[\ \mathbf{a}_\tau \otimes \mathbf{I}_M\ ]\,\mathbf{W}_\tau + \mathbf{S}_\tau = \underline{\mathbf{C}\mathbf{W}}_\tau\mathbf{a}_\tau + \mathbf{S}_\tau \ , \tag{12}$$

where $\underline{\mathbf{C}\mathbf{W}}_\tau = [\mathbf{C}_{\tau,1}\mathbf{W}_\tau\ldots\mathbf{C}_{\tau,p}\mathbf{W}_\tau]$. We augment the observation model in (12) with the first-order Markov models of $\mathbf{W}_\tau$ [6] and $\mathbf{a}_\tau$,

$$\begin{aligned}\mathbf{W}_\tau &= A\cdot\mathbf{W}_{\tau-1} + \Delta\mathbf{W}_\tau \ , \\ \mathbf{a}_\tau &= B\cdot\mathbf{a}_{\tau-1} + \Delta\mathbf{a}_\tau \ ,\end{aligned} \tag{13}$$

with $0 < \{A, B\} < 1$ as the respective transition coefficients. Our Markov modeling of the nonlinear expansion coefficients $\mathbf{a}_\tau$ shall facilitate the resulting algorithm to endure a time-varying nonlinearity due to temperature drifts, and to adapt to component tolerances [3]. For derivational ease, the process noise terms, i.e., $\Delta\mathbf{W}_\tau$ and $\Delta\mathbf{a}_\tau$, are modeled as zero-mean and bin-wise uncorrelated Gaussian random vectors with diagonal covariance matrices $\boldsymbol{\Psi}_\tau^{\Delta w} = \langle\Delta\mathbf{W}_\tau\Delta\mathbf{W}_\tau^{\mathrm{H}}\rangle$ and $\boldsymbol{\Psi}_\tau^{\Delta a} = \langle\Delta\mathbf{a}_\tau\Delta\mathbf{a}_\tau^{\mathrm{H}}\rangle$. The *composite* state-space model described by (12) and (13) is depicted in Fig. 2.

3. THE VARIATIONAL ALGORITHM: VB-SSFDAF

For joint learning of the composite state-space model, we seek to derive posterior estimators for the acoustic echo path and the non-

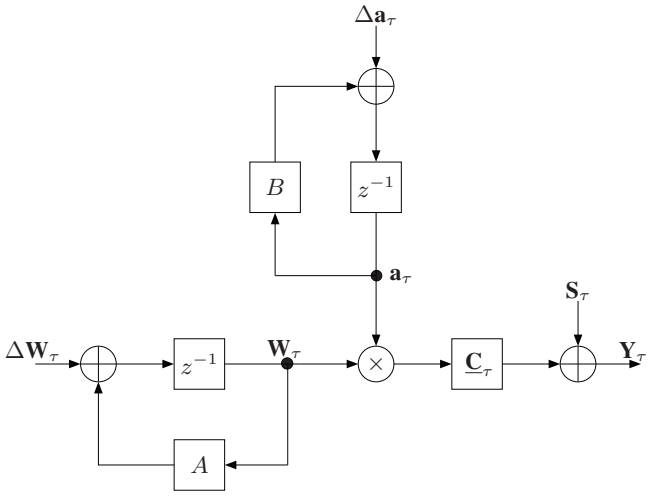

Fig. 2. Conceptual depiction of the dynamical modeling for the acoustic echo path and nonlinear expansion coefficients.

linear expansion coefficients. Since the true posterior $p(\mathbf{W}_\tau, \mathbf{a}_\tau|\mathbf{Y}_\tau)$ is generally intractable [5], we aim to estimate an arbitrary posterior distribution $q(\mathbf{W}_\tau, \mathbf{a}_\tau)$ on the echo path and nonlinear expansion coefficients that maximizes the *variational lower bound* for the log-evidence distribution $\ln\, p(\mathbf{Y}_\tau)$. For derivational simplicity we utilize the mean field approximation [5] to factorize the arbitrary posterior, i.e., $q(\mathbf{W}_\tau, \mathbf{a}_\tau) = q_w(\mathbf{W}_\tau)q_a(\mathbf{a}_\tau)$, and formulate the variational lower bound $\mathcal{F}(q_w(\mathbf{W}_\tau), q_a(\mathbf{a}_\tau))$ as [5]

$$\begin{aligned}\ln p(\mathbf{Y}_\tau) &\geq \int\int q_w(\mathbf{W}_\tau)q_a(\mathbf{a}_\tau)\ln\left(\frac{p(\mathbf{W}_\tau, \mathbf{a}_\tau, \mathbf{Y}_\tau)}{q_w(\mathbf{W}_\tau)q_a(\mathbf{a}_\tau)}\right) d\mathbf{W}_\tau d\mathbf{a}_\tau \\ &= \mathcal{F}(q_w(\mathbf{W}_\tau), q_a(\mathbf{a}_\tau))\ . \end{aligned} \tag{14}$$

Optimal log distributions, i.e., $\ln\, q_w(\mathbf{W}_\tau)$ and $\ln\, q_a(\mathbf{a}_\tau)$, that maximize $\mathcal{F}(q_w(\mathbf{W}_\tau), q_a(\mathbf{a}_\tau))$ to fit $\ln\, p(\mathbf{Y}_\tau)$ are then determined via the calculus of variations to be [5]

$$\ln\, q_w(\mathbf{W}_\tau) = \langle \ln\, p(\mathbf{W}_\tau, \mathbf{Y}_\tau|\mathbf{a}_\tau)\rangle_{q_a(\mathbf{a}_\tau)} + \ln\, Z_w\ , \tag{15}$$

$$\ln\, q_a(\mathbf{a}_\tau) = \langle \ln\, p(\mathbf{a}_\tau, \mathbf{Y}_\tau|\mathbf{W}_\tau)\rangle_{q_w(\mathbf{W}_\tau)} + \ln\, Z_a\ . \tag{16}$$

The operators $\langle\cdot\rangle_{q_a(\mathbf{a}_\tau)}$ and $\langle\cdot\rangle_{q_w(\mathbf{W}_\tau)}$ denote expectations with respect to $q_a(\mathbf{a}_\tau)$ and $q_w(\mathbf{W}_\tau)$, with Z_a and Z_w as the corresponding normalizations. The expressions in (15) and (16) signify the learning of one quantity of interest subject to the expectation over the other one. Due to Markov modeling, all distributions are implicitly conditioned on the predicted mean of the echo path $\widehat{\mathbf{W}}^+_{\tau-1}$ and the nonlinear coefficients $\widehat{\mathbf{a}}^+_{\tau-1}$ inferred at frame-time τ.

3.1. Echo-Path State Estimator

The joint distribution $p(\mathbf{W}_\tau, \mathbf{Y}_\tau|\mathbf{a}_\tau) = p(\mathbf{Y}_\tau|\mathbf{W}_\tau, \mathbf{a}_\tau)p(\mathbf{W}_\tau|\widehat{\mathbf{W}}^+_{\tau-1})$ in (15) is factorized in terms of complex multivariate Gaussian transmission [7] $p(\mathbf{Y}_\tau|\mathbf{W}_\tau, \mathbf{a}_\tau) = \mathcal{N}_c(\mathbf{G}\widetilde{\mathbf{X}}_\tau\mathbf{W}_\tau, \mathbf{\Psi}^S_\tau)$, i.e., according to (9), and transition $p(\mathbf{W}_\tau|\widehat{\mathbf{W}}^+_{\tau-1}) = \mathcal{N}_c(\widehat{\mathbf{W}}^+_{\tau-1}, \mathbf{P}^+_{\tau-1})$ distributions. The terms $\widehat{\mathbf{W}}^+_{\tau-1} = A\cdot\widehat{\mathbf{W}}_{\tau-1}$ and $\mathbf{P}^+_{\tau-1} = A^2\cdot\mathbf{P}_{\tau-1} + \mathbf{\Psi}^{\Delta w}_\tau$ [8] are the predicted echo path mean and the predicted error covariance at frame-time τ, respectively. Here, we highlight that $\widehat{\mathbf{W}}_{\tau-1}$ is the estimated DFT-domain echo path vector at frame-time $\tau-1$ and $\mathbf{P}_{\tau-1} = \left\langle(\mathbf{W}_{\tau-1} - \widehat{\mathbf{W}}_{\tau-1})(\mathbf{W}_{\tau-1} - \widehat{\mathbf{W}}_{\tau-1})^{\mathrm{H}}\right\rangle$ is the corresponding echo-path state-error covariance. Considering this factorization, we meticulously resolve the expectation [7] $\langle\cdot\rangle_{q_a(\mathbf{a}_\tau)}$ in (15). The subsequent completion of squares results in Kalman-like update rules for the acoustic echo path estimation, i.e., estimation of the mean and the covariance of $q_w(\mathbf{W}_\tau)$. We thus present the echo path learning rules in the diagonalized [6] [7] form

$$\widehat{\mathbf{W}}^+_{\tau-1} = A\cdot\widehat{\mathbf{W}}_{\tau-1} \tag{17}$$

$$\mathbf{P}^+_{\tau-1} = A^2\cdot\mathbf{P}_{\tau-1} + \mathbf{\Psi}^{\Delta w}_\tau \tag{18}$$

$$\mathbf{\Omega}_\tau = \frac{R}{M}\sum_{i=1}^{p}\mathbf{Q}_{\tau-1,ii}\mathbf{X}^{\mathrm{H}}_{\tau,i}{\mathbf{\Psi}^S_\tau}^{-1}\mathbf{X}_{\tau,i} \tag{19}$$

$$\widetilde{\mathbf{P}}^+_{\tau-1} = \left(\mathbf{\Omega}_\tau + {\mathbf{P}^+_{\tau-1}}^{-1}\right)^{-1}\ ;\ \mathbf{V}_\tau = \widetilde{\mathbf{P}}^+_{\tau-1}{\mathbf{P}^+_{\tau-1}}^{-1} \tag{20}$$

$$\mathbf{K}_{w,\tau} = \widetilde{\mathbf{P}}^+_{\tau-1}\widehat{\widetilde{\mathbf{X}}}^{\mathrm{H}}_\tau\left(\widehat{\widetilde{\mathbf{X}}}_\tau\widetilde{\mathbf{P}}^+_{\tau-1}\widehat{\widetilde{\mathbf{X}}}^{\mathrm{H}}_\tau + \frac{M}{R}\mathbf{\Psi}^S_\tau\right)^{-1} \tag{21}$$

$$\widehat{\mathbf{W}}_\tau = \mathbf{V}_\tau\widehat{\mathbf{W}}^+_{\tau-1} + \mathbf{K}_{w,\tau}\left(\mathbf{Y}_\tau - \mathbf{G}\widehat{\widetilde{\mathbf{X}}}_\tau\mathbf{V}_\tau\widehat{\mathbf{W}}^+_{\tau-1}\right) \tag{22}$$

$$\mathbf{P}_\tau = \left(\mathbf{I}_M - \frac{R}{M}\mathbf{K}_{w,\tau}\widehat{\widetilde{\mathbf{X}}}_\tau\right)\widetilde{\mathbf{P}}^+_{\tau-1}\ , \tag{23}$$

where $\widehat{\widetilde{\mathbf{X}}}_\tau = \sum_{i=1}^{p}\widehat{a}_{\tau-1,i}\mathbf{X}_{\tau,i}$. The computation of $\widehat{\widetilde{\mathbf{X}}}_\tau$ and the expression in (19) manifest the injection and utilization of the estimated nonlinear coefficients $\widehat{a}_{\tau-1,i}$ and the $p\times p$ coefficient error covariance $\mathbf{Q}_{\tau-1} = \left\langle(\mathbf{a}_{\tau-1} - \widehat{\mathbf{a}}_{\tau-1})(\mathbf{a}_{\tau-1} - \widehat{\mathbf{a}}_{\tau-1})^{\mathrm{H}}\right\rangle$ from the iteration at $\tau-1$. Owing to the modeled diagonal attributes of $\mathbf{Q}_{\tau-1}$, we have only considered the corresponding diagonal elements $\mathbf{Q}_{\tau-1,ii}$. We have initialized the echo-path state estimator with $\widehat{a}_{\tau-1,i}$ and $\mathbf{Q}_{\tau-1,ii}$, which were estimated at frame-time $\tau-1$, to eventually execute our variational algorithm in a purely sequential way.

If we set the coefficient error covariance to $\mathbf{Q}_{\tau-1}\to 0$, it causes the echo-path absorption term in (19) to $\mathbf{\Omega}_\tau\to 0$. Consequently, the modified predicted state-error covariance $\widetilde{\mathbf{P}}^+_{\tau-1}$ in (20) and the modified Kalman gain $\mathbf{K}_{w,\tau}$ in (21) acquire their conventional forms according to [7], and the echo-path assimilation term $\mathbf{V}_\tau$ equates to the identity matrix $\mathbf{I}_M$. Hence, for $\mathbf{Q}_{\tau-1}\to 0$ the recursion in (17) to (23) reduces to the SSFDAF [7] for a linear dynamical model.

3.2. Nonlinear Expansion Coefficient Estimator

We follow an approach similar to the derivation of the echo-path state estimator in Sec. 3.1 and factorize the joint distribution in (16). The distribution $p(\mathbf{a}_\tau, \mathbf{Y}_\tau|\mathbf{W}_\tau) = p(\mathbf{Y}_\tau|\mathbf{W}_\tau, \mathbf{a}_\tau)p(\mathbf{a}_\tau|\widehat{\mathbf{a}}^+_{\tau-1})$ can be expressed in terms of complex multivariate Gaussian transmission $p(\mathbf{Y}_\tau|\mathbf{W}_\tau, \mathbf{a}_\tau) = \mathcal{N}_c(\underline{\mathbf{CW}}_\tau\mathbf{a}_\tau, \mathbf{\Psi}^S_\tau)$, i.e., according to (12), and transition $p(\mathbf{a}_\tau|\widehat{\mathbf{a}}^+_{\tau-1}) = \mathcal{N}_c(\widehat{\mathbf{a}}^+_{\tau-1}, \mathbf{Q}^+_{\tau-1})$ distributions. The terms $\widehat{\mathbf{a}}^+_{\tau-1} = B\cdot\widehat{\mathbf{a}}_{\tau-1}$ and $\mathbf{Q}^+_{\tau-1} = B^2\cdot\mathbf{Q}_{\tau-1} + \mathbf{\Psi}^{\Delta a}_\tau$ [8] are the predicted nonlinear coefficient mean and the predicted coefficient error covariance at frame-time τ, respectively. Considering these representations, we resolve the expectation [7] $\langle\cdot\rangle_{q_w(\mathbf{W}_\tau)}$ in (16) and complete the squares to obtain the Kalman-like learning rules for the nonlinear expansion coefficient posterior

$$\widehat{\mathbf{a}}^+_{\tau-1} = B\cdot\widehat{\mathbf{a}}_{\tau-1} \tag{24}$$

$$\mathbf{Q}^+_{\tau-1} = B^2\cdot\mathbf{Q}_{\tau-1} + \mathbf{\Psi}^{\Delta a}_\tau \tag{25}$$

$$\mathbf{\Lambda}_{\tau,ii} = \frac{R}{M}\mathrm{Tr}\left\{\mathbf{X}^{\mathrm{H}}_{\tau,i}{\mathbf{\Psi}^S_\tau}^{-1}\mathbf{X}_{\tau,i}\mathbf{P}_\tau\right\} \tag{26}$$

$$\widetilde{\mathbf{Q}}^+_{\tau-1} = \left(\mathbf{\Lambda}_\tau + {\mathbf{Q}^+_{\tau-1}}^{-1}\right)^{-1}\ ;\ \mathbf{U}_\tau = \widetilde{\mathbf{Q}}^+_{\tau-1}{\mathbf{Q}^+_{\tau-1}}^{-1} \tag{27}$$

$$\mathbf{K}_{a,\tau} = \widetilde{\mathbf{Q}}^+_{\tau-1}\widehat{\underline{\mathbf{CW}}}^{\mathrm{H}}_\tau\left(\widehat{\underline{\mathbf{CW}}}_\tau\widetilde{\mathbf{Q}}^+_{\tau-1}\widehat{\underline{\mathbf{CW}}}^{\mathrm{H}}_\tau + \mathbf{\Psi}^S_\tau\right)^{-1} \tag{28}$$

$$\widehat{\mathbf{a}}_\tau = \mathbf{U}_\tau\widehat{\mathbf{a}}^+_{\tau-1} + \mathbf{K}_{a,\tau}\left(\mathbf{Y}_\tau - \widehat{\underline{\mathbf{CW}}}_\tau\mathbf{U}_\tau\widehat{\mathbf{a}}^+_{\tau-1}\right) \tag{29}$$

$$\mathbf{Q}_\tau = \left(\mathbf{I}_p - \mathbf{K}_{a,\tau}\widehat{\underline{\mathbf{CW}}}_\tau\right)\widetilde{\mathbf{Q}}^+_{\tau-1}\ , \tag{30}$$

where $\mathrm{Tr}\{\cdot\}$ denotes the trace operation, $\mathbf{K}_{a,\tau}$ is the modified Kalman gain for the nonlinear coefficient estimator, $\widehat{\underline{\mathbf{CW}}}_\tau = \left[\mathbf{C}_{\tau,1}\widehat{\mathbf{W}}_\tau\ldots\mathbf{C}_{\tau,p}\widehat{\mathbf{W}}_\tau\right]$, $\widetilde{\mathbf{Q}}^+_{\tau-1}$ is the modified predicted error covariance, and $\mathbf{U}_\tau$ is the nonlinear coefficient assimilation term.

With the exception of (29), we use $\mathbf{C}_{\tau,i}\widehat{\mathbf{W}}_\tau = \frac{R}{M}\mathbf{X}_{\tau,i}\widehat{\mathbf{W}}_\tau$ [6] [7] to achieve an efficient implementation. We consider only the main diagonals while computing the matrix-inverse in (28), and updating the terms in (26) and (30) for ensuring numerical stability of the estimator. It can be noticed in the computation of the nonlinear coefficient absorption term $\mathbf{\Lambda}_{\tau,ii}$ in (26) and in the described computation of $\widehat{\underline{\mathbf{CW}}}_\tau$ that the echo-path state-error covariance $\mathbf{P}_\tau$ and the echo-path state estimate $\widehat{\mathbf{W}}_\tau$ are being injected into the nonlinear coefficient estimator. Thus, it is evident that both of the partial posterior estimators are mutually coupled via the corresponding estimated mean and error covariances. We term the recursion in (24) to (30) together with the rules given in (17) to (23) as the VB-SSFDAF.

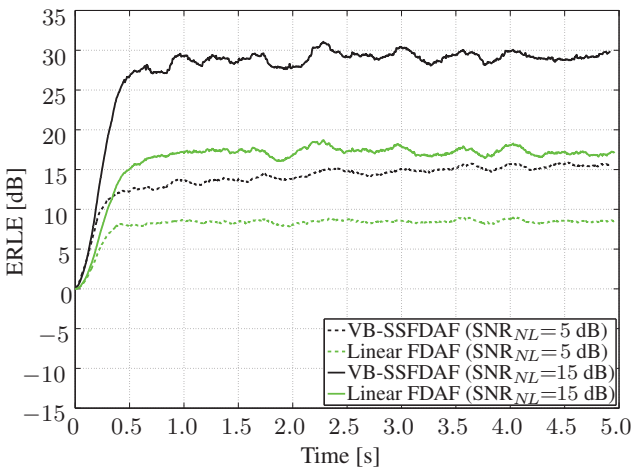

Fig. 3. Performance of the VB-SSFDAF and the linear FDAF at $\mathrm{SNR}_{NL} = 5$ dB and $\mathrm{SNR}_{NL} = 15$ dB, with ESR $= 60$ dB.

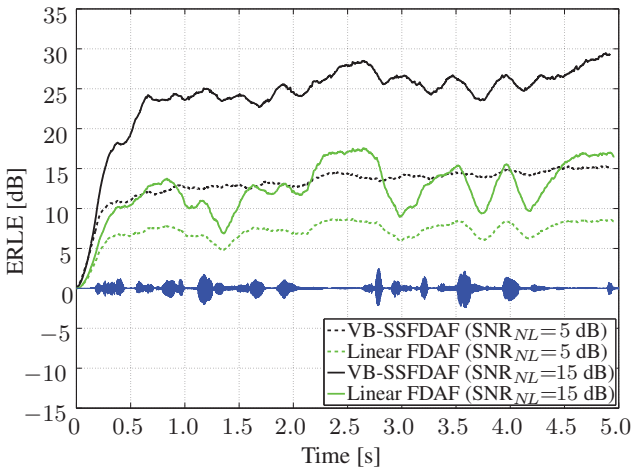

Fig. 4. Performance in a double-talk scenario at $\mathrm{SNR}_{NL} = 5$ dB and $\mathrm{SNR}_{NL} = 15$ dB, with ESR $= 0$ dB.

4. RESULTS

Hard clipping [3] was considered for simulating the nonlinear behavior of the loudspeaker. The clipping threshold of the hard clipping function was adjusted to achieve nonlinear signal-to-noise ratio $\mathrm{SNR}_{NL} = 10\log_{10}(\sigma^2_{x_t}/\sigma^2_{f(x_t)-x_t})$ of 5 dB and 15 dB, for the considered Gaussian input x_t. We configured $R = 64$ and $M = 256$ at the sampling frequency $f_s = 16$ kHz, with $A{=}B{=}0.9997$. The odd power series, i.e., $f(x_t) = \sum_{i=1}^{p} a'_{t,i} x_t^{2i-1}$, was used to model the underlying nonlinearity with the order $p = 5$. We set $\mathbf{Q}_{11,\tau} = 0$ and $a_{\tau,1} = 1$ so that the VB-SSFDAF may perform at least as good as a linear algorithm. The linear frequency-domain adaptive filter (FDAF) [9] was selected as the reference algorithm, while echo return loss enhancement $\mathrm{ERLE} = 10\log_{10}(\sigma^2_{d_t}/\sigma^2_{\widehat{d}_t - d_t})$ served as the instrumental measure of performance. The linear FDAF can be described via the respective error and update equations, i.e.,

$$\mathbf{E}_\tau = \mathbf{Y}_\tau - \mathbf{C}_{\tau,1}\widehat{\mathbf{W}}_{\tau-1} \tag{31}$$

$$\widehat{\mathbf{W}}_\tau = \widehat{\mathbf{W}}_{\tau-1} + \alpha {\boldsymbol{\Psi}_\tau^{\mathbf{X}_1}}^{-1} \mathbf{X}^{\mathrm{H}}_{\tau,1}\mathbf{E}_\tau \ , \tag{32}$$

where $\alpha = 0.1$ was the selected adaptation constant. A forgetting factor $\gamma = 0.9$ was used to recursively estimate the diagonalized DFT-domain input signal covariance matrix $\boldsymbol{\Psi}_\tau^{\mathbf{X}_1}$ [9]. The covariance parameters, i.e., $\boldsymbol{\Psi}_\tau^{S}$, $\boldsymbol{\Psi}_\tau^{\Delta w}$, and $\boldsymbol{\Psi}_\tau^{\Delta a}$, were estimated according to the expectation-maximization-type approach described in [7].

In Fig. 3, we compare the performance of the VB-SSFDAF with the linear FDAF. The echo-to-signal ratio $\mathrm{ESR} = 10\log_{10}(\sigma^2_{d_t}/\sigma^2_{s_t})$ was intentionally kept at a favorable ESR $= 60$ dB to first maintain the focus on the effects of nonlinear system modeling. In both cases of SNR_{NL} studied here, the VB-SSFDAF nearly doubles the attained ERLE as compared to the linear FDAF. We further investigate the robustness of the derived algorithm in the presence of severe double-talk at ESR $= 0$ dB, using the near-end speech signal presented in Fig. 4. We observe that the VB-SSFDAF, as compared to the linear FDAF, continues to adapt robustly despite severe double-talk and considerable system nonlinearity.

5. CONCLUSIONS

We have presented a DFT-domain algorithm for nonlinear acoustic echo cancellation that is derived using the variational Bayesian methodology. The acoustic echo path and the unknown expansion coefficients of the memoryless loudspeaker nonlinearity have been modeled as random variables with first-order Markov property. The variational Bayesian solution, i.e., VB-SSFDAF, for the formulated composite state-space model then comprises Kalman-like posterior estimators for the acoustic echo path and the nonlinear expansion coefficients. The derived posterior estimators can be implemented efficiently and perform robust echo cancellation even in the presence of harsh nonlinearity and continuous double-talk.

6. REFERENCES

[1] F. Küch, A. Mitnacht, and W. Kellermann, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, Philadelphia, PA, Mar. 2005, vol. 3, pp. 105–108.

[2] S. Malik and G. Enzner, "Fourier expansion of Hammerstein models for nonlinear acoustic system identification," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, Prague, CZ, May 2011, pp. 85–88.

[3] A. Stenger and W. Kellermann, "Adaptation of memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Process.*, vol. 80, no. 9, pp. 1747–1760, Sep. 2000.

[4] M. I. Mossi, C. Yemdji, N. Evans, C. Beaugeant, and P. Degry, "Robust and low-cost cascade non-linear acoustic echo cancellation," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, Prague, CZ, May 2011, pp. 89–92.

[5] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, New York, 1st ed. 2006. corr. 2nd printing edition, Oct. 2007.

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

[7] S. Malik and G. Enzner, "Online maximum-likelihood learning of time-varying dynamical models in block-frequency domain," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, Dallas, TX, Mar. 2010, pp. 3822–3825.

[8] L. L. Scharf, *Statistical Signal Processing*, Addison-Wesley, Reading, MA, 1991.

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