Online Estimation of Reverberation Parameters For Late Residual Echo Suppression

Naveen Kumar Desiraju, Simon Doclo, *Senior Member, IEEE*, Markus Buck, *Member, IEEE*, and Tobias Wolff

***Abstract*—In hands-free telephony and other distant-talk applications, often a short AEC filter is used to achieve fast convergence at low computational cost. As a result, a significant amount of late residual echo (LRE) may remain, especially in highly reverberant environments. This LRE can be suppressed using a postfilter in the subband domain, which requires an estimate of the power spectral density (PSD) of the LRE. To estimate the LRE PSD, an exponentially decaying model with frequency-dependent reverberation scaling and decay parameters has frequently been assumed. State-of-the-art methods estimate both reverberation parameters independently of each other, either in offline or in online mode. In this article, we propose two signal-based methods (i.e. output error and equation error) to jointly estimate both reverberation parameters in online mode. The estimated parameters are then used to generate an estimate for the LRE PSD, which is fed into a postfilter for the purpose of late residual echo suppression. We derive several gradient-descent-based algorithms to simultaneously update both reverberation parameters, minimizing either the mean squared error or the mean squared log error cost function. The proposed methods are compared with state-of-the-art methods in terms of the accuracy of the estimated reverberation parameters and the corresponding LRE PSD estimate. Extensive simulation results using both artificial as well as measured room impulse responses show that the proposed output error method with mean squared log error minimization outperforms state-of-the-art methods in all considered scenarios.**

***Index Terms*—Acoustic echo cancellation, adaptive filters, late residual echo estimation, residual echo suppression.**

I. Introduction

HANDS-FREE telephony and other distant-talk applications, such as voice-controlled multimedia devices, are often used in large reverberant rooms, where the distance between the desired (near-end) speaker and the microphone may be quite large. Due to the acoustic coupling between the loudspeaker and the microphone, the microphone signal is typically degraded by the acoustic echo of the far-end signal, which may significantly reduce the quality and/or the intelligibility of the near-end speaker. Acoustic echo cancellation (AEC) [1] is a key technology used in such scenarios, aimed at canceling the echo from the microphone signal. An AEC system typically consists of an adaptive filter [2], [3] which estimates the acoustic echo path, i.e. the room impulse response (RIR) between the loudspeaker and the microphone. The adaptive filter is used to generate an estimate of the acoustic echo signal, which is subsequently subtracted from the microphone signal. The resulting signal is referred to as the AEC error signal and is composed of near-end speech, background noise and usually some residual echo, as the AEC filter is unable to completely accurately estimate the RIR in practice (filter misalignment). When deploying an AEC system in a room with a large reverberation time (T_{60}), a large filter length needs to be used in order to achieve good echo cancellation performance. However, using a long filter results in large computational cost for updating the filter and may also lead to slow filter convergence [2], [3]. Hence, aiming at achieving fast filter convergence at low computational cost, in practice often a short AEC filter is used, which however results in a large amount of late residual echo (LRE).

In practice, a postfilter is often used in addition to the AEC filter, aimed at suppressing the residual echo and background noise while not distorting the near-end speech signal. Although multi-frame postfilters have been proposed [4], most postfilters are single-tap real-valued gains [5]–[12]. To design the postfilter in the subband domain, an accurate estimate of the power spectral density (PSD) of the residual echo and background noise signals is required. A simple but frequently used method to estimate the PSD of the residual echo signal is to apply a coupling factor to the far-end signal PSD, where the coupling factor is estimated during periods of near-end speech absence [1]. However, since this method does not take into account any temporal context and is unable to model the LRE PSD accurately, its performance is quite poor, especially when using a short AEC filter. Hence, several other LRE PSD estimators have been proposed which are based on the statistical reverberation model proposed in [13], [14], which assumes that the late reverberant part of a RIR decays exponentially at a rate proportional to the T_{60}. These PSD estimators require estimates of two parameters: the reverberation decay parameter (corresponding to the T_{60}) and the reverberation scaling parameter (a.k.a. initial power of the LRE).

Manuscript received February 19, 2019; revised August 22, 2019; accepted October 13, 2019. Date of publication October 21, 2019; date of current version December 24, 2019. This work was supported by the European Union's Seventh Framework Programme (FP7/2007-2013) project DREAMS under Grant ITN-GA-2012-316969. The associate editor coordinating the review of this manuscript and approving it for publication was M. de Diego. *(Corresponding author: Naveen Kumar Desiraju.)*

N. K. Desiraju was with Acoustic Speech Enhancement Research, Nuance Communications Deutschland GmbH, 89077 Ulm, Germany. He is currently with Harman Connected Services GmbH, 85748 Garching, Germany (e-mail: naveen.desiraju@harman.com).

S. Doclo is with the Department of Medical Physics and Acoustics and the Cluster of Excellence Hearing4All, University of Oldenburg, 26111 Oldenburg, Germany (e-mail: simon.doclo@uni-oldenburg.de).

M. Buck and T. Wolff were with Acoustic Speech Enhancement Research, Nuance Communications Deutschland GmbH, 89077 Ulm, Germany. They are currently with Cerence Inc., 89077 Ulm, Germany (e-mail: markus.buck@cerence.com; tobias.wolff@cerence.com).

Digital Object Identifier 10.1109/TASLP.2019.2948765

To estimate both reverberation parameters, channel-based as well as signal-based methods have been proposed. *Channel-based* methods [12], [15] estimate the reverberation parameters using the coefficients of the converged AEC filter, either assuming frequency-dependent [12] or frequency-independent parameters [15]. Channel-based methods are only effective if relatively long AEC filters are used, which are able to capture the decay of the late reverberant part of the RIR. *Signal-based* methods, on the other hand, estimate the reverberation parameters directly from the far-end and the residual echo signals [16]–[18]. In [16], a signal-based method was proposed to estimate both reverberation parameters independently of each other in *offline* mode (i.e. batch processing). The reverberation scaling parameter was estimated by minimizing the mean squared error (MSE) cost function, while the reverberation decay parameter was estimated by minimizing the mean squared log error (MSLE) cost function. In [17], a pure acoustic echo suppression system, i.e. without an AEC filter, was considered and a recursive estimator for the (residual) echo PSD was used. A signal-based method exploiting higher-order-statistics was proposed to estimate the initial power of the (residual) echo and the reverberation decay parameter independently of each other in *online mode*. Using the recursive estimator for the LRE PSD in [12], in [18] we proposed two signal-based methods, namely an output error and an equation error method, to *jointly* estimate both reverberation parameters in offline mode. These methods, which were originally proposed to estimate the coefficients of generic IIR filters in the time-domain [2], [19], [20], were applied on PSDs to jointly estimate both reverberation parameters by minimizing either the MSE or the MSLE cost function.

Based on the work in [18], in this paper we propose methods to jointly estimate both reverberation parameters in *online* mode. The estimated parameters are then used to generate an estimate for the LRE PSD, which is fed into a postfilter for the purpose of late residual echo suppression. We derive several gradient-descent-based algorithms to simultaneously update both parameters, minimizing either the MSE or the MSLE cost function. In particular, we propose to use the recursive prediction error (RPE) and pseudo-linear regression (PLR) algorithms, which were derived for time-domain recursive systems [19], to update the parameters for the output error method. The different signal-based methods (output/equation error), algorithms (RPE/PLR) and cost functions (MSE/MSLE) are compared with state-of-the-art signal-based methods [16], [17] in terms of accuracy of the reverberation parameter estimates and the corresponding LRE PSD estimate, and in terms of the resulting residual echo suppression and near-end speech distortion.

The paper is organized as follows. The signal model as well as some basic AEC and postfiltering principles are presented in Section II. The recursive estimator for the LRE PSD, the different proposed signal-based parameter estimation methods and the gradient-descent-based algorithms to simultaneously update both parameters are presented in Sections III, IV and V, respectively. Section VI presents the simulation results comparing the performance of the proposed methods with state-of-the-art methods using both artificially generated as well as measured RIRs.

II. Signal Model and AEC System

Fig. 1 shows a loudspeaker-enclosure-microphone (LEM) system with the far-end signal x, the acoustic echo signal d, the near-end speech signal s, the background noise signal v and the microphone signal y. The RIR characterizing the acoustic echo path between the loudspeaker and the microphone is denoted as h and assumed to be time-invariant and of length N_h. The microphone signal at discrete-time sample n is given as:

$$y(n) = s(n) + v(n) + \underbrace{\sum\nolimits_{i=0}^{N_h-1} h(i) \cdot x(n-i)}_{d(n)}. \qquad (1)$$

For the subband processing, a fast Fourier transform (FFT) filterbank of order N_{FFT} is used to transform the (windowed) time-domain signals into the short-time Fourier transform (STFT) domain, with the total number of subbands given by $K = \frac{N_{\text{FFT}}}{2} + 1$. The complex-valued STFT coefficients of the far-end signal x in subband k and frame ℓ are computed as:

$$X(k,\ell) = \sum_{m=0}^{N_{\text{FFT}}-1} x(\ell \cdot F + m) \cdot W_{\text{ana}}(m) \cdot e^{-j\frac{2\pi}{N_{\text{FFT}}}km}, \qquad (2)$$

where $j = \sqrt{-1}$, F denotes the frameshift and W_{ana} denotes the analysis window. Similarly to (2), the STFT coefficients of $s(n)$, $v(n)$, $d(n)$ and $y(n)$ are denoted as $S(k,\ell)$, $V(k,\ell)$, $D(k,\ell)$ and $Y(k,\ell)$, respectively.

The complete AEC system consists of two components: an (adaptive) AEC filter estimating the echo path and a residual echo suppression (RES) postfilter. Both components will be explained in more detail in the following subsections.

A. Acoustic Echo Cancellation

To cancel the acoustic echo signal from the microphone signal, we consider a G-tap subband AEC filter $\hat{\underline{H}}$. The acoustic echo estimate is given as:

$$\hat{D}(k,\ell) = \underline{X}^H(k,\ell)\, \hat{\underline{H}}(k), \qquad (3)$$

with

$$\underline{X}(k,\ell) = \left[\, X(k,\ell)\ \ldots\ X(k,\ell-G+1) \,\right]^T \qquad (4)$$

the G-dimensional tap-input vector to the AEC filter $\hat{\underline{H}}$:

$$\hat{\underline{H}}(k) = \left[\, \hat{H}_1(k)\ \ldots\ \hat{H}_G(k) \,\right]^T, \qquad (5)$$

where $\cdot^H$ denotes the Hermitian operator and $\cdot^T$ denotes the transpose operator.

The signal obtained after the acoustic echo estimate is subtracted from the microphone signal is referred to as the AEC error signal:

$$\begin{aligned} E(k,\ell) &= Y(k,\ell) - \hat{D}(k,\ell) \\ &= S(k,\ell) + V(k,\ell) + \underbrace{\Big(D(k,\ell) - \hat{D}(k,\ell)\Big)}_{R(k,\ell)}, \end{aligned} \qquad (6)$$

where R denotes the residual echo signal, which consists of the early residual echo signal R_E (due to filter misalignment)

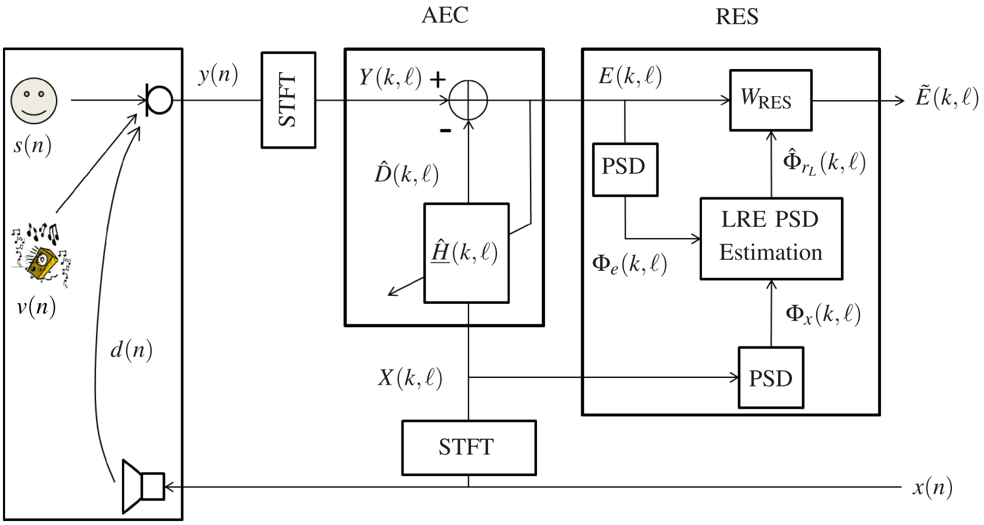

Fig. 1. Acoustic echo cancellation (AEC) and residual echo suppression (RES) systems.

and the LRE signal R_L (due to the limited length of the AEC filter). In this paper, the filter length G is chosen so as to cover the direct path and the early reflections of the RIR h, i.e. $G = \lfloor \frac{N}{F} \rfloor$, where N corresponds to the length of the direct path and early reflections in samples. This means that the LRE signal R_L is assumed to contain only late reverberation. Additionally, we assume no filter misalignment, i.e. $R_E = 0$, such that the residual echo signal only consists of the late residual echo signal, i.e. $R = R_L$.

B. Residual Echo Suppression

Residual echo suppression can be performed in the subband domain by applying a real-valued gain W_{RES} to the AEC error signal E, as shown in Fig. 1. A frequently used gain is the Wiener filter [1], which is derived by assuming that the signals S, R_L and V are independent stationary stochastic processes, leading to:

$$W_{\text{RES}}(k,\ell) = 1 - \frac{\lambda_{r_L}(k,\ell) + \lambda_v(k,\ell)}{\lambda_e(k,\ell)}. \tag{7}$$

Here, λ_{r_L}, λ_v and λ_e denote the PSDs of the LRE, the background noise and the AEC error signals, respectively, defined as $\lambda_{r_L}(k,\ell) = \mathscr{E}\{|R_L(k,\ell)|^2\}$, $\lambda_v(k,\ell) = \mathscr{E}\{|V(k,\ell)|^2\}$, and $\lambda_e(k,\ell) = \mathscr{E}\{|E(k,\ell)|^2\}$, where $\mathscr{E}\{\cdot\}$ denotes the statistical expectation operator. In practice, the statistical expectation operator is approximated by temporal averaging (assuming ergodicity), e.g.:

$$\Phi_e(k,\ell) = \alpha \cdot \Phi_e(k,\ell-1) + (1-\alpha) \cdot |E(k,\ell)|^2, \tag{8}$$

where Φ_e is an approximation of the PSD λ_e and α denotes the smoothing factor. The quantities Φ_{r_L}, Φ_v and Φ_x are defined similarly as in (8) and are approximations of λ_{r_L}, λ_v and λ_x, respectively. Please note that for an unobservable signal such as r_L, the quantity Φ_{r_L} itself needs to be estimated, with the estimate denoted as $\hat{\Phi}_{r_L}$. In the remainder of the paper, we will use the term *true PSD* to refer to λ_a, $a \in \{e, r_L, v, x\}$, the term *PSD* to refer to its approximation Φ_a, $a \in \{e, r_L, v, x\}$ and the term *PSD estimate* to refer to its estimate for an unobservable signal $\hat{\Phi}_a$, $a \in \{r_L, v\}$.

In order to control the aggressiveness of the residual echo suppression, we use the following gain for the RES postfilter:

$$\boxed{W_{\text{RES}}(k,\ell) = \max\left\{1 - \beta \cdot \left(\frac{\hat{\Phi}_{r_L}(k,\ell) + \hat{\Phi}_v(k,\ell)}{\Phi_e(k,\ell)}\right), \gamma\right\}} \tag{9}$$

with over-estimation factor β and spectral floor γ. While the AEC error PSD Φ_e is directly observable, the LRE PSD Φ_{r_L} and the background noise PSD Φ_v need to be estimated. Many approaches have been proposed in literature for estimating the background noise PSD [21]–[23]. In this paper, we assume that the background noise is stationary and its PSD is known.

The processed AEC error signal is given as:

$$\tilde{E}(k,\ell) = W_{\text{RES}}(k,\ell) \cdot E(k,\ell), \tag{10}$$

which can be expressed as the sum of its individual components in a similar way to (6):

$$\tilde{E}(k,\ell) = \tilde{S}(k,\ell) + \tilde{V}(k,\ell) + \tilde{R}_L(k,\ell), \tag{11}$$

where $\tilde{S}$, $\tilde{V}$ and $\tilde{R}_L$ are obtained by multiplying S, V and R_L with the RES postfilter, similarly to (10). For the purpose of evaluation, these processed signals are then synthesized to the time-domain using inverse STFT and overlap-add processing, yielding the time-domain signals $\tilde{e}(n)$, $\tilde{s}(n)$, $\tilde{v}(n)$ and $\tilde{r}_L(n)$, respectively.

III. Model for LRE PSD

In [13], an exponentially decaying model for the late reverberant part of a RIR was proposed when the source-microphone distance is larger than the critical distance, defined as the distance where the energy of the direct sound is equal to the energy of all reflections [24]. According to this model, the late reverberant part of a RIR can be described as a realization of a stochastic

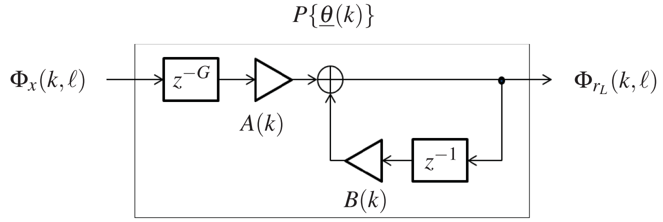

Fig. 2. Model for LRE PSD Φ_{r_L} as a function of far-end signal PSD Φ_x.

process:

$$h(i) = w_L(i) \cdot e^{-\rho\,(i-N)}, \quad N \leq i < N_h, \tag{12}$$

where N_h denotes the total length of the RIR in samples, w_L is a zero-mean white Gaussian noise process with variance σ_L^2 and ρ denotes the decay rate. The decay rate is related to the T_{60} as:

$$\rho = \frac{3 \cdot \ln 10}{f_s \cdot T_{60}}, \tag{13}$$

where f_s denotes the sampling frequency in Hz. Although in (13) it is assumed that the T_{60} is frequency-independent, it should be noted that in practice the T_{60} (and hence the decay rate ρ) is frequency-dependent [24].

As mentioned in Section II-A, we assume that the AEC filter is able to cancel the direct sound component and the early reflections, such that the LRE signal R_L contains only late reverberation. Based on the RIR model in (12), a recursive expression for λ_{r_L} can be derived (see Appendix A), i.e.:

$$\lambda_{r_L}(k,\ell) = A \cdot \lambda_x(k,\ell-G) + B \cdot \lambda_{r_L}(k,\ell-1), \tag{14}$$

where A denotes the reverberation scaling parameter and B denotes the reverberation decay parameter. These parameters are related to the parameters σ_L^2 and ρ of the RIR model in (12) as (see Appendix A):

$$A = \sigma_L^2 \cdot \left(\frac{1-e^{-2\rho F}}{1-e^{-2\rho}}\right), \tag{15}$$

$$B = e^{-2\rho F}. \tag{16}$$

In this paper, we assume the reverberation parameters to be *frequency-dependent*, such that similarly to (14), a recursive expression for Φ_{r_L} using frequency-dependent parameters can be obtained as in [12]:

$$\boxed{\Phi_{r_L}(k,\ell) = A(k) \cdot \Phi_x(k,\ell-G) + B(k) \cdot \Phi_{r_L}(k,\ell-1),} \tag{17}$$

with the parameters $A(k)$ and $B(k)$ given as:

$$A(k) = \sigma_L^2(k) \cdot \left(\frac{1-e^{-2\rho(k)F}}{1-e^{-2\rho(k)}}\right), \tag{18}$$

$$B(k) = e^{-2\rho(k)F}. \tag{19}$$

The expression in (17) relating the LRE PSD Φ_{r_L} to the far-end signal PSD Φ_x is illustrated in Fig. 2 using the IIR filter $P\{\underline{\theta}(k)\}$, where

$$\underline{\theta}(k) = \left[A(k)\ B(k)\right]^T. \tag{20}$$

In the next section, we will present different methods to estimate $\underline{\theta}(k)$. It should be noted that $\underline{\theta}(k)$ is estimated during periods of near-end speech absence and subsequently used to estimate the LRE PSD during periods of double-talk.

IV. Parameter Estimation Methods

Several methods have been proposed in literature to estimate both reverberation parameters A and B independently of each other. In [12], a channel-based method was proposed using the converged AEC filter coefficients. In [16], a signal-based method was proposed in offline mode (i.e. batch processing), where the parameter A was estimated by minimizing an MSE cost function and the parameter B was estimated by minimizing an MSLE cost function. In [17], an acoustic echo suppression setup without an AEC filter (i.e. $G = 0$) was considered and a signal-based method based on higher-order statistics was proposed to estimate both parameters in online mode. For the purpose of fair comparison, we consider a slightly modified version of the method in [17] in order to estimate the LRE PSD for our considered setup and compare this method with our proposed parameter estimation methods (see Section VI). Since we assume a perfect AEC filter (see Section II-A), this modification simply corresponds to inserting a delay of G frames in the original method in [17] (details presented in Appendix B).

To *jointly* estimate the parameters of generic IIR filters in the time-domain, several signal-based methods have been proposed [2], [19], [20], [25]–[27], either based on the output error (OE) or the equation error (EE). In [18], we applied the OE and EE methods on PSDs to jointly estimate both reverberation parameters in offline mode (i.e. batch processing), minimizing either the MSE or the MSLE cost function. Simulation results showed that the most accurate estimates for the reverberation decay parameter B and the LRE PSD Φ_{r_L} were obtained using the OE method minimizing the MSLE cost function, while the most accurate estimates for the reverberation scaling parameter A were obtained using either the OE or the EE method minimizing the MSE cost function.

Based on the offline methods from [18], in this paper we investigate the OE and EE methods in *online* mode to jointly estimate both reverberation parameters A and B during periods of near-end speech absence, where the parameters are simultaneously updated in each frame using a gradient-descent-based algorithm (see Section V). The estimated parameters $\hat{\underline{\theta}}(k)$ are then fed into the IIR filter $P\{\hat{\underline{\theta}}(k,\ell)\}$ to estimate the LRE PSD (also during double-talk), as illustrated in Fig. 3:

$$\boxed{\hat{\Phi}_{r_L}(k,\ell) = \hat{A}(k,\ell) \cdot \Phi_x(k,\ell-G) + \hat{B}(k,\ell) \cdot \hat{\Phi}_{r_L}(k,\ell-1).} \tag{21}$$

In the following subsections we will discuss the OE and EE methods to estimate the reverberation parameters $\hat{\underline{\theta}}(k)$.

A. Output Error Method

The OE method is a well-known method used for parameter estimation of linear recursive systems in a variety of applications. The OE method is characterized by the following *recursive*

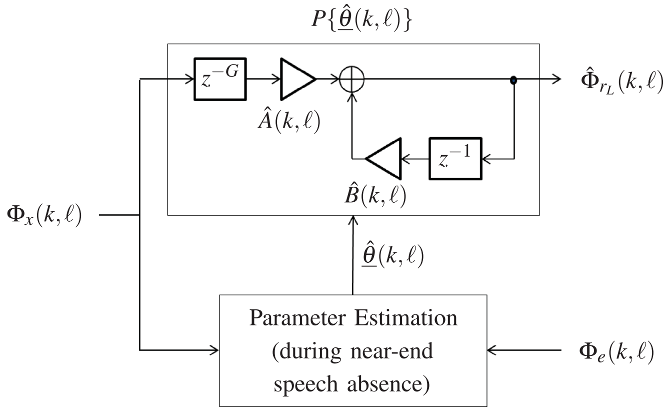
$P\{\hat{\underline{\theta}}(k,\ell)\}$
z^{-G}
$\hat{A}(k,\ell)$
$\hat{\Phi}_{r_L}(k,\ell)$
z^{-1}
$\hat{B}(k,\ell)$
$\Phi_x(k,\ell)$
$\hat{\underline{\theta}}(k,\ell)$
Parameter Estimation (during near-end speech absence)
$\Phi_e(k,\ell)$

Fig. 3. The LRE PSD estimate $\hat{\Phi}_{rL}$ is computed using the far-end signal PSD Φ_x, with the parameters $\hat{\underline{\theta}}(k,\ell)$ estimated during near-end speech absence.

difference equation (where the superscript $^{\mathrm{O}}$ denotes the OE method):

$$\hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell) = \hat{A}^{\mathrm{O}}(k,\ell)\cdot\Phi_x(k,\ell-G) + \hat{B}^{\mathrm{O}}(k,\ell)\cdot\hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1), \tag{22}$$

with the corresponding IIR filter structure illustrated in Fig. 4. Here, $\hat{\Phi}^{\mathrm{O}}_{r_L}$ denotes the OE PSD estimate and

$$\hat{\underline{\theta}}^{\mathrm{O}}(k,\ell) = \left[\hat{A}^{\mathrm{O}}(k,\ell)\ \hat{B}^{\mathrm{O}}(k,\ell)\right]^T \tag{23}$$

denotes the reverberation parameters estimated using the OE method, which are fed into (21) to generate the LRE PSD estimate $\hat{\Phi}_{r_L}$. Please note that (22) has the same recursive structure as (21), such that $\hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell) = \hat{\Phi}_{r_L}(k,\ell)$. From (22), it can be observed that the OE PSD estimate in the current frame $\hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell)$ not only depends on the parameter estimates in the current frame $\hat{\underline{\theta}}^{\mathrm{O}}(k,\ell)$, but also on the OE PSD estimate in the previous frame $\hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1)$, which itself depends on the parameter estimates in the previous frame $\hat{\underline{\theta}}^{\mathrm{O}}(k,\ell-1)$, and so on. Thus, $\hat{\Phi}^{\mathrm{O}}_{r_L}$ is a *non-linear* function of $\hat{\underline{\theta}}^{\mathrm{O}}$, where the current OE PSD estimate depends on the parameter estimates in all previous frames.

The output error is obtained by subtracting the output in (22) from the target PSD Φ_{r_L}:

$$Q^{\mathrm{O}}(k,\ell) = \Phi_{r_L}(k,\ell) - \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell). \tag{24}$$

Similarly, the output log error is given as:

$$Q^{\mathrm{O}}_{\ln}(k,\ell) = \ln\,\Phi_{r_L}(k,\ell) - \ln\,\hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell) = \ln\left(\frac{\Phi_{r_L}(k,\ell)}{\hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell)}\right). \tag{25}$$

To compute the parameter estimates, we will consider minimizing either the MSE or the MSLE cost function:

$$\mathscr{J}^{\mathrm{O}}_{\mathrm{MSE}}\left(\hat{A}^{\mathrm{O}}(k,\ell),\hat{B}^{\mathrm{O}}(k,\ell)\right) = \mathscr{E}\left\{\left[Q^{\mathrm{O}}(k,\ell)\right]^2\right\}, \tag{26}$$

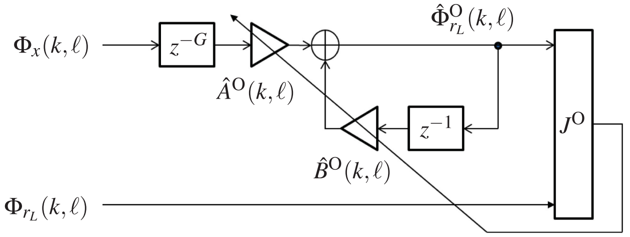
$\Phi_x(k,\ell)$
z^{-G}
$\hat{A}^{\mathrm{O}}(k,\ell)$
$\hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell)$
z^{-1}
$\hat{B}^{\mathrm{O}}(k,\ell)$
J^{O}
$\Phi_{r_L}(k,\ell)$

Fig. 4. Parameter estimation using the output error method by minimizing the cost function J^{O}.

$$\mathscr{J}^{\mathrm{O}}_{\mathrm{MSLE}}\left(\ln\hat{A}^{\mathrm{O}}(k,\ell),\ln\hat{B}^{\mathrm{O}}(k,\ell)\right) = \mathscr{E}\left\{\left[Q^{\mathrm{O}}_{\ln}(k,\ell)\right]^2\right\}. \tag{27}$$

To update the parameters in every frame using a gradient-descent-based algorithm (see Section V), these cost functions will be approximated by their instantaneous values:

$$J^{\mathrm{O}}_{\mathrm{MSE}}\left(\hat{A}^{\mathrm{O}}(k,\ell),\hat{B}^{\mathrm{O}}(k,\ell)\right) = \left[Q^{\mathrm{O}}(k,\ell)\right]^2 = \left[\Phi_{r_L}(k,\ell) - \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell)\right]^2, \tag{28}$$

$$J^{\mathrm{O}}_{\mathrm{MSLE}}\left(\ln\hat{A}^{\mathrm{O}}(k,\ell),\ln\hat{B}^{\mathrm{O}}(k,\ell)\right) = \left[Q^{\mathrm{O}}_{\ln}(k,\ell)\right]^2 = \left[\ln\left(\frac{\Phi_{r_L}(k,\ell)}{\hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell)}\right)\right]^2. \tag{29}$$

As $\hat{\Phi}^{\mathrm{O}}_{r_L}$ is a non-linear function of the parameters $\hat{\underline{\theta}}^{\mathrm{O}}$, the cost functions $J^{\mathrm{O}}_{\mathrm{MSE}}$ and $J^{\mathrm{O}}_{\mathrm{MSLE}}$ are not quadratic in the parameters and may exhibit multiple local minima [19], [27]–[30]. This may result in gradient-descent-based algorithms converging to a local minimum, thereby yielding sub-optimal and inaccurate parameter estimates, with the initial value of $\hat{\underline{\theta}}^{\mathrm{O}}$ also influencing to which minimum the algorithms converge. This is a typical problem when using adaptive IIR filters for identifying recursive systems [19].

B. Equation Error Method

In order to avoid the local minima problem associated with the OE method, the EE method has often been employed for parameter estimation of linear recursive systems [19], [20]. The EE method differs from the OE method by using the delayed target PSD $\Phi_{r_L}(k,\ell-1)$ instead of the delayed PSD estimate $\hat{\Phi}_{r_L}(k,\ell-1)$ for computing the current PSD estimate, thereby breaking the recursive structure. The EE method is characterized by the following *non-recursive* difference equation (where the superscript $^{\mathrm{E}}$ denotes the EE method):

$$\hat{\Phi}^{\mathrm{E}}_{r_L}(k,\ell) = \hat{A}^{\mathrm{E}}(k,\ell)\cdot\Phi_x(k,\ell-G) + \hat{B}^{\mathrm{E}}(k,\ell)\cdot\Phi_{r_L}(k,\ell-1), \tag{30}$$

with the corresponding non-recursive filter structure illustrated in Fig. 5. Here, $\hat{\Phi}^{\mathrm{E}}_{r_L}$ denotes the EE PSD estimate and

$$\hat{\underline{\theta}}^{\mathrm{E}}(k,\ell) = \left[\hat{A}^{\mathrm{E}}(k,\ell)\quad\hat{B}^{\mathrm{E}}(k,\ell)\right]^T \tag{31}$$

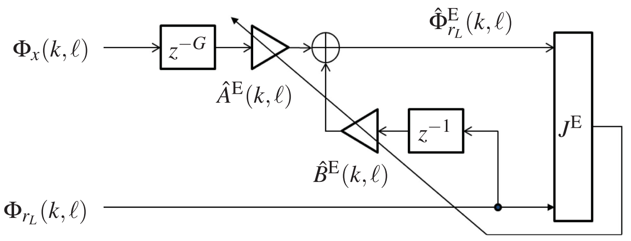
$\Phi_x(k,\ell)$
z^{-G}
$\hat{A}^{\mathrm{E}}(k,\ell)$
$\hat{\Phi}^{\mathrm{E}}_{r_L}(k,\ell)$
z^{-1}
J^{E}
$\hat{B}^{\mathrm{E}}(k,\ell)$
$\Phi_{r_L}(k,\ell)$

Fig. 5. Parameter estimation using the equation error method by minimizing the cost function J^{E}.

denotes the reverberation parameters estimated using the EE method, which are fed into (21) to generate the LRE PSD estimate $\hat{\Phi}_{r_L}$. As a result, the PSD estimate $\hat{\Phi}^{\mathrm{E}}_{r_L}$ is a *linear* function of $\underline{\hat{\theta}}^{\mathrm{E}}$. Please note that using $\Phi_{r_L}(k,\ell-1)$ instead of $\hat{\Phi}_{r_L}(k,\ell-1)$ in (30) is an approximation, such that unlike the OE method, the EE PSD estimate $\hat{\Phi}^{\mathrm{E}}_{r_L}$ is not equal to the LRE PSD estimate $\hat{\Phi}_{r_L}$.

Similarly to (24), the equation error is given as:

$$Q^{\mathrm{E}}(k,\ell) = \Phi_{r_L}(k,\ell) - \hat{\Phi}^{\mathrm{E}}_{r_L}(k,\ell), \tag{32}$$

and similarly to (25), the equation log error is given as:

$$Q^{\mathrm{E}}_{\ln}(k,\ell) = \ln\,\Phi_{r_L}(k,\ell) - \ln\,\hat{\Phi}^{\mathrm{E}}_{r_L}(k,\ell) = \ln\left(\frac{\Phi_{r_L}(k,\ell)}{\hat{\Phi}^{\mathrm{E}}_{r_L}(k,\ell)}\right). \tag{33}$$

Contrary to the cost functions $J^{\mathrm{O}}_{\mathrm{MSE}}$ and $J^{\mathrm{O}}_{\mathrm{MSLE}}$, the cost functions $J^{\mathrm{E}}_{\mathrm{MSE}}$ and $J^{\mathrm{E}}_{\mathrm{MSLE}}$ (defined similarly to (28) and (29) respectively) are quadratic in the parameters, hence exhibiting a single global minimum and no local minima [19], [20]. This makes the EE method particularly attractive for use in practical applications, as the corresponding adaptive algorithms typically have fast convergence and converge to a global minimum. However, it has been shown in [19] that the EE method yields biased solutions in the presence of additive noise, where the bias is proportional to the amount of noise. Additionally, as Φ_x and Φ_{r_L} are approximations of the true PSDs λ_x and λ_{r_L} (see Section II-B), these approximations introduce additional noise to the system. This results in the EE method yielding biased solutions even in the absence of additive noise, as was observed in [18] when using the EE method for reverberation parameter estimation in offline mode. In this paper, we investigate how accurately the EE method estimates the reverberation parameters in online mode.

V. Gradient-Descent-Based Algorithms

In this section, we derive gradient-descent-based algorithms to update the reverberation parameters $\underline{\theta}(k)$ in every frame for the OE and EE estimation methods, either minimizing the MSE or the MSLE cost function.

For both estimation methods, the gradient-descent update rule for the MSE cost function is given as:

$$\boxed{\underline{\hat{\theta}}^{\mathrm{I}}(k,\ell+1) = \underline{\hat{\theta}}^{\mathrm{I}}(k,\ell) - \frac{\Gamma}{2}\odot\underline{\nabla}^{\mathrm{I}}_{\mathrm{MSE}}(k,\ell),} \tag{34}$$

where $\mathrm{I}\in\{\mathrm{O},\mathrm{E}\}$ denotes the used estimation method, $\odot$ denotes element-wise multiplication, $\underline{\Gamma} = [\,\mu_A\ \mu_B\,]^T$ denotes the (fixed) step-sizes used update both parameters, and

$$\underline{\nabla}^{\mathrm{I}}_{\mathrm{MSE}}(k,\ell) = \left[\frac{\partial J^{\mathrm{I}}_{\mathrm{MSE}}\left(\hat{A}^{\mathrm{I}}(k,\ell),\hat{B}^{\mathrm{I}}(k,\ell)\right)}{\partial \hat{A}^{\mathrm{I}}(k,\ell)}\ \frac{\partial J^{\mathrm{I}}_{\mathrm{MSE}}\left(\hat{A}^{\mathrm{I}}(k,\ell),\hat{B}^{\mathrm{I}}(k,\ell)\right)}{\partial \hat{B}^{\mathrm{I}}(k,\ell)}\right]^T \tag{35}$$

denotes the gradient of the MSE cost function. Using (28), the partial derivatives of the MSE cost function with respect to the reverberation scaling and decay parameter estimates are equal to:

$$\frac{\partial J^{\mathrm{I}}_{\mathrm{MSE}}\left(\hat{A}^{\mathrm{I}}(k,\ell),\hat{B}^{\mathrm{I}}(k,\ell)\right)}{\partial \hat{A}^{\mathrm{I}}(k,\ell)} = -2\cdot Q^{\mathrm{I}}(k,\ell)\cdot\frac{\partial\hat{\Phi}^{\mathrm{I}}_{r_L}(k,\ell)}{\partial \hat{A}^{\mathrm{I}}(k,\ell)}, \tag{36}$$

$$\frac{\partial J^{\mathrm{I}}_{\mathrm{MSE}}\left(\hat{A}^{\mathrm{I}}(k,\ell),\hat{B}^{\mathrm{I}}(k,\ell)\right)}{\partial \hat{B}^{\mathrm{I}}(k,\ell)} = -2\cdot Q^{\mathrm{I}}(k,\ell)\cdot\frac{\partial\hat{\Phi}^{\mathrm{I}}_{r_L}(k,\ell)}{\partial \hat{B}^{\mathrm{I}}(k,\ell)}. \tag{37}$$

The partial derivatives of the LRE PSD estimate $\hat{\Phi}^{\mathrm{I}}_{r_L}$ with respect to the parameter estimates will be computed for the OE and EE methods in subsections V-A and V-B, respectively. It should be noted that when minimizing the MSE cost function, the parameter updates in each frame depend on the error Q^{I} between the LRE PSD and its estimate.

For both estimation methods, the gradient-descent update rule for the MSLE cost function is given in the logarithmic domain[1] as:

$$\boxed{\ln\underline{\hat{\theta}}^{\mathrm{I}}(k,\ell+1) = \ln\underline{\hat{\theta}}^{\mathrm{I}}(k,\ell) - \frac{\Gamma}{2}\odot\underline{\nabla}^{\mathrm{I}}_{\mathrm{MSLE}}(k,\ell),} \tag{38}$$

where the gradient of the MSLE cost function $\underline{\nabla}^{\mathrm{I}}_{\mathrm{MSLE}}$ is composed of the partial derivatives of the MSLE cost function with respect to the logarithm of the parameter estimates:

$$\underline{\nabla}^{\mathrm{I}}_{\mathrm{MSLE}}(k,\ell) = \left[\frac{\partial J^{\mathrm{I}}_{\mathrm{MSLE}}\left(\ln\hat{A}^{\mathrm{I}}(k,\ell),\ln\hat{B}^{\mathrm{I}}(k,\ell)\right)}{\partial\ln\hat{A}^{\mathrm{I}}(k,\ell)}\right.$$
$$\left.\frac{\partial J^{\mathrm{I}}_{\mathrm{MSLE}}\left(\ln\hat{A}^{\mathrm{I}}(k,\ell),\ln\hat{B}^{\mathrm{I}}(k,\ell)\right)}{\partial\ln\hat{B}^{\mathrm{I}}(k,\ell)}\right]^T. \tag{39}$$

Using (29), these partial derivatives are equal to:

$$\frac{\partial J^{\mathrm{I}}_{\mathrm{MSLE}}\left(\ln\hat{A}^{\mathrm{I}}(k,\ell),\ln\hat{B}^{\mathrm{I}}(k,\ell)\right)}{\partial\ln\hat{A}^{\mathrm{I}}(k,\ell)} = -2\cdot\left[\frac{Q^{\mathrm{I}}_{\ln}(k,\ell)}{\hat{\Phi}^{\mathrm{I}}_{r_L}(k,\ell)}\right]\cdot\frac{\partial\hat{\Phi}^{\mathrm{I}}_{r_L}(k,\ell)}{\partial\ln\hat{A}^{\mathrm{I}}(k,\ell)}, \tag{40}$$

$$\frac{\partial J^{\mathrm{I}}_{\mathrm{MSLE}}\left(\ln\hat{A}^{\mathrm{I}}(k,\ell),\ln\hat{B}^{\mathrm{I}}(k,\ell)\right)}{\partial\ln\hat{B}^{\mathrm{I}}(k,\ell)} = -2\cdot\left[\frac{Q^{\mathrm{I}}_{\ln}(k,\ell)}{\hat{\Phi}^{\mathrm{I}}_{r_L}(k,\ell)}\right]\cdot\frac{\partial\hat{\Phi}^{\mathrm{I}}_{r_L}(k,\ell)}{\partial\ln\hat{B}^{\mathrm{I}}(k,\ell)}. \tag{41}$$

[1] It should be noted that the gradient-descent update rule for the MSLE cost function in the linear domain yielded unreliable results.

The partial derivatives of the LRE PSD estimate $\hat{\Phi}^{\mathrm{I}}_{r_L}$ with respect to the logarithm of the parameter estimates will be computed for the OE and EE methods in subsections V-A and V-B, respectively. It should be noted that when minimizing the MSLE cost function, the parameter updates in each frame are normalized by the LRE PSD estimate $\hat{\Phi}^{\mathrm{I}}_{r_L}$ and depend on the log error $Q^{\mathrm{I}}_{\ln}$, which in turn depends on the ratio of the LRE PSD and its estimate.

A. Algorithms for Output Error method

Using (22), the partial derivatives of $\hat{\Phi}^{\mathrm{O}}_{r_L}$ with respect to the parameter estimates are equal to:

$$\begin{aligned}\frac{\partial \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell)}{\partial \hat{A}^{\mathrm{O}}(k,\ell)} &= \Phi_x(k,\ell-G) + \hat{B}^{\mathrm{O}}(k,\ell)\cdot\frac{\partial \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1)}{\partial \hat{A}^{\mathrm{O}}(k,\ell)},\\ \frac{\partial \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell)}{\partial \hat{B}^{\mathrm{O}}(k,\ell)} &= \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1) + \hat{B}^{\mathrm{O}}(k,\ell)\cdot\frac{\partial \hat{\Phi}^{O}_{r_L}(k,\ell-1)}{\partial \hat{B}^{O}(k,\ell)},\end{aligned} \tag{42}$$

while the partial derivatives of $\hat{\Phi}^{\mathrm{O}}_{r_L}$ with respect to the logarithm of the parameter estimates are equal to:

$$\begin{aligned}\frac{\partial \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell)}{\partial \ln \hat{A}^{\mathrm{O}}(k,\ell)} &= \hat{A}^{\mathrm{O}}(k,\ell)\quad\cdot \Phi_x(k,\ell-G)\\ &\quad + \hat{B}^{\mathrm{O}}(k,\ell)\cdot\frac{\partial \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1)}{\partial \ln \hat{A}^{\mathrm{O}}(k,\ell)},\\ \frac{\partial \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell)}{\partial \ln \hat{B}^{\mathrm{O}}(k,\ell)} &= \hat{B}^{\mathrm{O}}(k,\ell)\quad\cdot \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1)\\ &\quad + \hat{B}^{\mathrm{O}}(k,\ell)\cdot\frac{\partial \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1)}{\partial \ln \hat{B}^{\mathrm{O}}(k,\ell)}.\end{aligned} \tag{43}$$

It should be noted that (42) and (43) contain partial derivatives of the OE PSD estimate $\hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1)$ in the *previous* frame with respect to the parameter estimates $\hat{\underline{\theta}}^{\mathrm{O}}(k,\ell)$ and their logarithm $\ln\hat{\underline{\theta}}^{\mathrm{O}}(k,\ell)$ in the *current* frame, respectively. These terms appear due to the recursive filter structure of the OE method. These partial derivatives cannot be computed in a straightforward manner, as $\hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1)$ does not directly depend on $\hat{\underline{\theta}}^{\mathrm{O}}(k,\ell)$. In [19], two approximations have been proposed for computing these partial derivatives, which we now apply to the problem at hand.

1) Recursive Prediction Error (RPE): Although the OE PSD estimate $\hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1)$ in the previous frame does not directly depend on the parameter estimates $\hat{\underline{\theta}}^{\mathrm{O}}(k,\ell)$ in the current frame, it obviously directly depends on the parameter estimates $\hat{\underline{\theta}}^{\mathrm{O}}(k,\ell-1)$ in the previous frame. For computing the partial derivatives in (42) and (43), the RPE adaptive algorithm [19] uses the following approximations:

$$\begin{aligned}\frac{\partial \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1)}{\partial \hat{\underline{\theta}}^{\mathrm{O}}(k,\ell)} &\approx \frac{\partial \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1)}{\partial \hat{\underline{\theta}}^{\mathrm{O}}(k,\ell-1)},\\ \frac{\partial \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1)}{\partial \ln\hat{\underline{\theta}}^{\mathrm{O}}(k,\ell)} &\approx \frac{\partial \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1)}{\partial \ln\hat{\underline{\theta}}^{\mathrm{O}}(k,\ell-1)},\end{aligned} \tag{44}$$

which have been shown to be reasonable if the step-sizes $\underline{\Gamma}$ in (34) and (38) are sufficiently small. Using these approximations makes it possible to compute the partial derivatives in (42) and (43) recursively. As a result, both reverberation parameters are updated even when the respective inputs to the parameters are absent.

2) Pseudo Linear Regression (PLR): The PLR algorithm is an approximate gradient method [19] which assumes that the OE PSD estimate in the previous frame $\hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1)$ is independent of the parameter estimates in the current frame $\hat{\underline{\theta}}^{\mathrm{O}}(k,\ell)$, i.e.:

$$\begin{aligned}\frac{\partial \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1)}{\partial \hat{\underline{\theta}}^{\mathrm{O}}(k,\ell)} &= 0,\\ \frac{\partial \hat{\Phi}^{\mathrm{O}}_{r_L}(k,\ell-1)}{\partial \ln\hat{\underline{\theta}}^{\mathrm{O}}(k,\ell)} &= 0.\end{aligned} \tag{45}$$

Using (45) in (42) and (43) yields non-recursive formulations for the partial derivatives. It should be noted that the gradient computed using the PLR algorithm is an approximate version of the gradient computed using the RPE algorithm, as the assumptions in (45) are stronger than in (44).

B. Algorithm for Equation Error Method

Using (30), the partial derivatives of $\hat{\Phi}^{\mathrm{E}}_{r_L}$ with respect to the parameter estimates are equal to:

$$\begin{aligned}\frac{\partial \hat{\Phi}^{\mathrm{E}}_{r_L}(k,\ell)}{\partial \hat{A}^{\mathrm{E}}(k,\ell)} &= \Phi_x(k,\ell-G),\\ \frac{\partial \hat{\Phi}^{\mathrm{E}}_{r_L}(k,\ell)}{\partial \hat{B}^{\mathrm{E}}(k,\ell)} &= \Phi_{r_L}(k,\ell-1),\end{aligned} \tag{46}$$

while the partial derivatives of $\hat{\Phi}^{\mathrm{E}}_{r_L}$ with respect to the logarithm of the parameter estimates are equal to:

$$\begin{aligned}\frac{\partial \hat{\Phi}^{\mathrm{E}}_{r_L}(k,\ell)}{\partial \ln \hat{A}^{\mathrm{E}}(k,\ell)} &= \hat{A}^{\mathrm{E}}(k,\ell)\cdot\Phi_x(k,\ell-G),\\ \frac{\partial \hat{\Phi}^{\mathrm{E}}_{r_L}(k,\ell)}{\partial \ln \hat{B}^{\mathrm{E}}(k,\ell)} &= \hat{B}^{\mathrm{E}}(k,\ell)\cdot\Phi_{r_L}(k,\ell-1).\end{aligned} \tag{47}$$

Hence, the partial derivatives obtained for the EE method in (46) and (47) are non-recursive and similar to those obtained for the PLR algorithm for the OE method. It can also be observed that the reverberation parameters are not updated when the respective inputs to the parameters are absent, i.e. the reverberation scaling parameter $\hat{A}^{\mathrm{E}}$ is not updated when $\Phi_x(k,\ell-G)=0$, while the reverberation decay parameter $\hat{B}^{\mathrm{E}}$ is not updated when $\Phi_{r_L}(k,\ell-1)=0$.

TABLE I
NUMBER OF RIRS MEASURED IN EACH ROOM AND THE CORRESPONDING REVERBERATION TIMES (T_{60})

Room	**No. of RIRs**	$\mathbf{T_{60}}$
Lab	16	300-400 ms
Garage	16	400-500 ms
Office	16	500-600 ms
Echoic Room	7	850-950 ms

VI. SIMULATIONS

In this section, we evaluate the performance of the proposed online parameter estimation methods (OE and EE), cost functions (MSE and MSLE) and gradient-descent-based algorithms, giving rise to 6 combinations: OE-RPE-MSE, OE-PLR-MSE, EE-MSE, OE-RPE-MSLE, OE-PLR-MSLE and EE-MSLE. In Sections VI-A and VI-B we describe the signals and the algorithmic parameters used in our simulations, while in Section VI-C we discuss the performance metrics used to evaluate the PSD estimation accuracy, the residual echo suppression and the near-end speech distortion. In Section VI-D we perform two experiments to evaluate the performance of the proposed parameter estimation methods. To evaluate the parameter estimation accuracy, the first experiment is performed in an idealistic setting using artificial RIRs with frequency-independent reverberation parameters. The second experiment is performed in a realistic setting using RIRs measured in different rooms, comparing the performance of the proposed methods with state-of-the-art signal-based methods.

A. Signals

In our simulations, we use time-domain signals at a sampling frequency $f_s = 16$ kHz. The far-end speech signal x of length 30 secs and the near-end speech signal s of length 5 secs are obtained from the TIMIT database [31], where the double-talk condition occurs in the last 5 secs. The background noise signal v of length 30 secs is stationary air conditioner noise measured in an office. The time-domain signals are transformed into the STFT domain with $N_{\text{FFT}} = 512$ (i.e. $K = 257$) using a Hann analysis window and an overlap of 75%, i.e. a frameshift of $F = 128$.

The different RIRs used for our simulations can be divided into two categories:

- Artificial RIRs: A total of 30 RIRs were generated exactly according to the model in (12) with $N = 640$ and $N_h = 16000$ for all combinations of the frequency-independent parameters $\sigma_L^2 = \{-40, -36, -32, -28, -24, -20\}$ dB and $T_{60} = \{200, 400, 600, 800, 1000\}$ ms.
- Measured RIRs: A total of 55 RIRs were measured in 4 rooms with different reverberation times, with the number of RIRs measured in each room and the corresponding T_{60} values shown in Table I. The broadband T_{60} of each RIR was estimated by line-fitting on its corresponding energy decay curve [32]. The lab, garage and the echoic room were rectangular shaped, while the office room was L-shaped. It should be noted that these RIRs obviously don't exactly correspond to the model in (12).

TABLE II
STEP-SIZES USED FOR THE OE-RPE, OE-PLR AND EE METHODS (FOR BOTH THE MSE AND MSLE COST FUNCTIONS)

Method	MSLE		MSE	
	$\mu_{\mathbf{A}}$	$\mu_{\mathbf{B}}$	$\mu_{\mathbf{A}}$	$\mu_{\mathbf{B}}$
OE-RPE	10^{-2}	10^{-4}	10^{-4}	$10^{-3.5}$
OE-PLR	$10^{-1.75}$	$10^{-3.75}$	$10^{-2.5}$	10^{-2}
EE	10^{-1}	$10^{-2.5}$	10^{-2}	$10^{-1.5}$

B. Algorithmic Parameters

All required PSDs are computed via recursive smoothing according to (8), with the smoothing factor $\alpha = e^{\frac{-2 \cdot F}{f_s \cdot t_c}}$ computed for a time-constant $t_c = 0.02$ s. For the different combinations of parameter estimation methods, cost functions and gradient-descent-based algorithms, the step-sizes listed in Table II were used, which were found to give good results. In our experiments we however observed that the results obtained for the MSLE cost function were not very sensitive to the choice of the step-size. For the modified version of Favrot's method (see Appendix B), the delay M has been chosen as $M = N = 640$, while the delay P has been chosen as $P = \kappa \cdot F$ for two different values $\kappa = 12$ and $\kappa = 16$. In the RES postfilter in (9), an over-estimation factor $\beta = 2$ and a fixed spectral floor $\gamma = -20$ dB have been used.

C. Performance Metrics

To evaluate the accuracy of the LRE PSD estimate $\hat{\Phi}_{r_L}$, we compute the Log Spectral Distance (LSD) [23] between the PSD estimate and the target PSD Φ_{r_L}, which can be expressed as the sum of the under- and over-estimation scores:

$$\text{LSD} = \text{LSD}_{\text{un}} + \text{LSD}_{\text{ov}}, \tag{48}$$

$$\text{LSD}_{\text{un}} = \frac{10}{K \cdot L} \cdot \sum_{k=0}^{K-1} \sum_{\ell=l_1+1}^{l_1+L} \max\left\{0, \log_{10}\left(\frac{\Phi_{r_L}(k,\ell)}{\hat{\Phi}_{r_L}(k,\ell)}\right)\right\},$$

$$\text{LSD}_{\text{ov}} = -\frac{10}{K \cdot L} \cdot \sum_{k=0}^{K-1} \sum_{\ell=l_1+1}^{l_1+L} \min\left\{0, \log_{10}\left(\frac{\Phi_{r_L}(k,\ell)}{\hat{\Phi}_{r_L}(k,\ell)}\right)\right\},$$

where l_1 and L denote the start and the duration of the evaluation window (in frames), respectively. We choose l_1 corresponding to 20 secs ($l_1 = 2500$) and L corresponding to 5 secs ($L = 625$). A small LSD score corresponds to an accurate PSD estimate, with the perfect estimate $\hat{\Phi}_{r_L} = \Phi_{r_L}$ yielding LSD $= 0$.

To evaluate the amount of residual echo suppression and near-end speech distortion obtained by applying the RES postfilter, we compute the segmental residual echo attenuation (REA) and the segmental speech-to-speech distortion ratio (SSDR) [23], [33] respectively. The segmental REA is defined as:

$$\text{REA}_{\text{seg}} = \frac{1}{L} \cdot \sum_{\ell=l_1+1}^{l_1+L} \delta(\ell), \tag{49}$$

where

$$\delta(\ell) = 10 \cdot \log_{10}\left(\frac{\sum_{m=0}^{F-1} r_L^2(m + \ell \cdot F)}{\sum_{m=0}^{F-1} \tilde{r}_L^2(m + \ell \cdot F)}\right) \tag{50}$$

denotes the REA in each frame, with the late residual echo signal r_L and the processed residual echo signal $\tilde{r}_L$ obtained through inverse STFT processing of R_L and $\tilde{R}_L$ (see (11)), respectively. A large REA_{seg} means that a large amount of residual echo has been suppressed. The segmental SSDR is defined as:

$$\text{SSDR}_{\text{seg}} = \frac{1}{L} \cdot \sum_{\ell=l_2+1}^{l_2+L} \eta(\ell), \tag{51}$$

where

$$\eta(\ell) = 10 \cdot \log_{10} \left(\frac{\sum_{m=0}^{F-1} s^2(m+\ell \cdot F)}{\sum_{m=0}^{F-1} \Big(s(m+\ell \cdot F) - \tilde{s}(m+\ell \cdot F) \Big)^2} \right) \tag{52}$$

denotes the SSDR in each frame, with s the near-end speech signal and $\tilde{s}$ the processed near-end speech signal (see (11)). Here, we choose l_2 corresponding to 25 secs ($l_2 = 3125$), such that the segmental SSDR is computed in the last 5 secs when double-talk occurs. A large SSDR_{seg} corresponds to a small near-end speech signal distortion. In general, a trade-off exists between obtaining large residual echo attenuation and small near-end speech distortion. Hence, it is desirable to maximize REA_{seg} while keeping SSDR_{seg} as large as possible.

D. Experimental Results

The first experiment is performed in an idealistic setting, i.e. using artificial RIRs, a perfect AEC filter, no near-end speech and no background noise. In this experiment we evaluate how accurately the proposed methods estimate the RIR parameters and the LRE PSD. The second experiment is performed in a realistic setting using measured RIRs, a converged (but not perfect) subband AEC filter, near-end speech and background noise. In this experiment, we compare the LSD, segmental REA and SSDR scores and the T_{60} estimates obtained using the proposed online methods with those obtained using state-of-the-art methods, i.e. Valero's method [16] (offline version) and Favrot's method [17] (modified online version presented in Appendix B).

1) Idealistic Setting: As already mentioned, in this experiment we use artificial RIRs with frequency-independent parameters σ_L^2 and T_{60} (see Section VI-A) to generate the acoustic echo signal and we assume a perfect AEC filter, i.e. no early residual echo ($R_E = 0$). Additionally, we assume that no near-end speech and background noise are present, i.e. $s(n) = v(n) = 0$, such that $E(k,\ell) = R_L(k,\ell)$. For this idealistic setting, we compare the estimates of the RIR parameters $\hat{\sigma}_L^2$ and $\hat{T}_{60}$ with the true values, and compare the LSD scores of the LRE PSD estimates obtained using the OE-RPE, OE-PLR and EE methods (for both the MSE and MSLE cost functions). For each method, the parameter estimates $\hat{\sigma}_L^2$ and $\hat{T}_{60}$ are obtained by averaging the converged values of the estimated model parameters $A(k)$ and $B(k)$ over all frequency bins and using them in (15), (16) and (13).

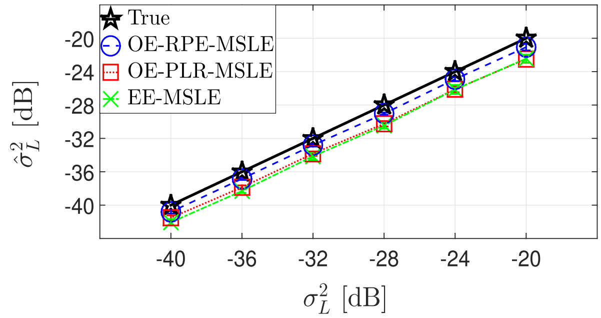

Fig. 6. Plot of $\hat{\sigma}_L^2$ vs σ_L^2 for the OE-RPE, OE-PLR and EE methods when minimizing the MSLE cost function for the idealistic setting.

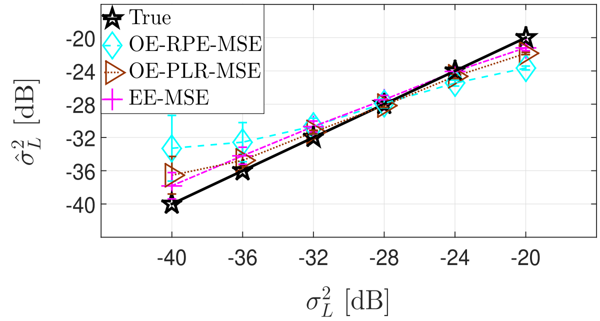

Fig. 7. Plot of $\hat{\sigma}_L^2$ vs σ_L^2 for the OE-RPE, OE-PLR and EE methods when minimizing the MSE cost function for the idealistic setting.

Fig. 6 and Fig. 7 show the estimated scaling parameter $\hat{\sigma}_L^2$ as a function of the true scaling parameter σ_L^2 for the OE-RPE, OE-PLR and EE methods when minimizing the MSLE and MSE cost functions, respectively. Each point in these figures corresponds to the average result obtained for 5 RIRs (with different T_{60} values), while the error bars depict the standard deviation across these 5 RIRs. On the one hand, it can be observed that for MSLE minimization (Fig. 6), all considered methods slightly underestimate σ_L^2 and yield a very small standard deviation, indicating robustness to different T_{60} values. On the other hand, for MSE minimization (Fig. 7), all considered methods yield less accurate estimates with large standard deviations. Overall, the OE-RPE method with MSLE minimization gives the most accurate results for all considered σ_L^2 and T_{60}.

Fig. 8 and Fig. 9 show the estimated reverberation time $\hat{T}_{60}$ as a function of the true reverberation time T_{60} for the OE-RPE, OE-PLR and EE methods when minimizing the MSLE and MSE cost functions, respectively. Each point in these figures now corresponds to the average result obtained for 6 RIRs (with different σ_L^2), while the error bars depict the standard deviation across these 6 RIRs. It can be observed that for MSLE minimization (Fig. 8), the OE-RPE method estimates the T_{60} very accurately, while the OE-PLR and EE methods slightly over-estimate the T_{60}. All three methods yield small standard deviations, indicating robustness to different σ_L^2 values. For the MSE minimization (Fig. 9), the OE-RPE and OE-PLR methods estimate the T_{60} reasonably accurately with large standard deviations, while the EE method fails completely, especially for large

TABLE III
AVERAGE LSD SCORES OBTAINED FOR ARTIFICIALLY GENERATED RIRS FOR ALL PROPOSED PARAMETER ESTIMATION METHODS

Method	LSD_{un}					LSD_{ov}				
	$T_{60} = 0.2$ s	0.4 s	0.6 s	0.8 s	1 s	$T_{60} = 0.2$ s	0.4 s	0.6 s	0.8 s	1 s
OE-RPE-MSLE	0.84	0.98	1.07	1.19	1.28	1.24	1.36	1.47	1.54	1.63
OE-PLR-MSLE	1.37	1.57	1.67	1.63	1.59	1.01	0.96	1.00	1.09	1.13
EE-MSLE	1.83	2.18	2.44	2.45	2.48	0.67	0.67	0.64	0.69	0.67
OE-RPE-MSE	1.15	0.90	0.95	0.99	1.02	1.25	1.97	2.51	2.93	3.10
OE-PLR-MSE	0.73	0.86	0.91	0.94	0.97	1.39	1.82	2.22	2.4	2.42
EE-MSE	0.81	0.41	0.18	0.08	0.06	1.38	2.9	5.27	7.66	9.01

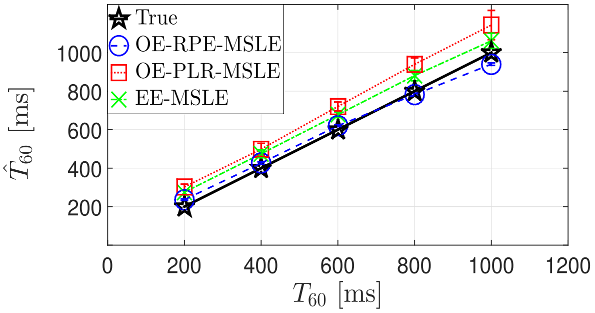

Fig. 8. Plot of $\hat{T}_{60}$ vs T_{60} for the OE-RPE, OE-PLR and EE methods when minimizing the MSLE cost function for the idealistic setting.

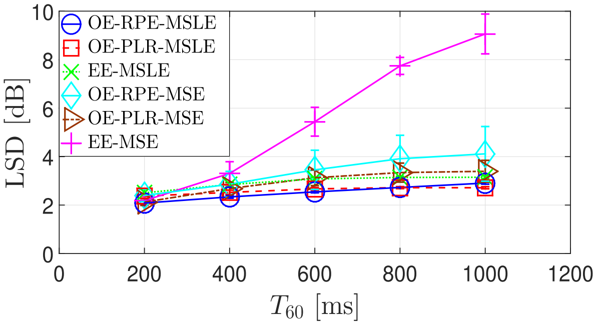

Fig. 10. Plot of LSD vs T_{60} for all proposed parameter estimation methods for the idealistic setting.

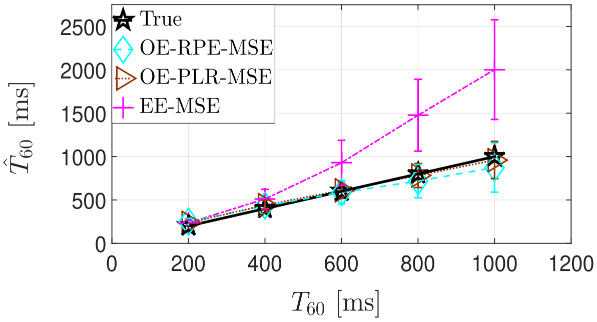

Fig. 9. Plot of $\hat{T}_{60}$ vs T_{60} for the OE-RPE, OE-PLR and EE methods when minimizing the MSE cost function for the idealistic setting.

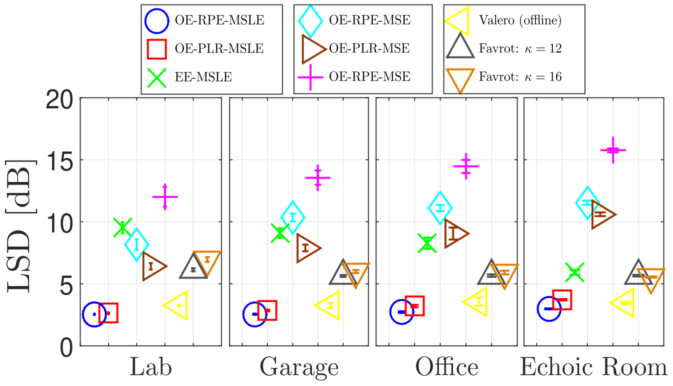

Fig. 11. LSD scores obtained for RIRs measured in four different rooms for all considered parameter estimation methods.

T_{60}. Overall, the OE-RPE method with MSLE minimization gives the most accurate and consistent results for all considered σ_L^2 and T_{60}.

Fig. 10 shows the LSD scores of the LRE PSD estimates obtained using all considered methods as a function of T_{60}. Each point in this figure again corresponds to the average result obtained for 6 RIRs (with different σ_L^2), while the error bars depict the standard deviation across these 6 RIRs. Additionally, Table III breaks down all average LSD scores into under- and over-estimation scores (see (48)). From these results it can be observed that the OE-RPE-MSLE and OE-PLR-MSLE methods consistently outperform all other methods across all T_{60} values, yielding the lowest LSD scores with the smallest standard deviations. When minimizing the MSE cost function, all methods yield significantly larger over-estimation scores than under-estimation scores, especially for large T_{60} values.

In conclusion, based on the results obtained for the idealistic setting, the OE-RPE-MSLE method outperforms all other proposed methods in terms of estimation accuracy of the RIR parameters σ_L^2 and T_{60} and the LRE PSD Φ_{r_L}. This corresponds to the result obtained in [18] for offline processing.

2) Realistic Setting: In this experiment, we use measured RIRs (see Table I) to generate the acoustic echo signal and subband AEC filter to perform echo cancellation (see Section II-A). For the AEC filter we have used a rather short filter length ($G = 5$ frames, corresponding to 64 ms), aiming at canceling the direct sound component and the early reflections, while achieving fast convergence at low computational cost. The subband filter was pre-converged using the NLMS algorithm [2], with white Gaussian noise as the far-end signal. It should be noted that when using a subband AEC filter, the early residual echo is

TABLE IV
AVERAGE LSD SCORES OBTAINED FOR RIRS MEASURED IN FOUR ROOMS (SEE TABLE I) FOR ALL CONSIDERED PARAMETER ESTIMATION METHODS

Method	Lab		Garage		Office		Echoic Room	
	LSD$_{un}$	LSD$_{ov}$	LSD$_{un}$	LSD$_{ov}$	LSD$_{un}$	LSD$_{ov}$	LSD$_{un}$	LSD$_{ov}$
OE-RPE-MSLE	1.00	1.54	1.03	1.52	1.17	1.58	1.40	1.63
OE-PLR-MSLE	1.30	1.33	1.26	1.59	1.90	1.30	1.45	2.27
EE-MSLE	0.23	9.29	0.20	8.89	0.28	7.99	0.49	5.43
OE-RPE-MSE	0.43	7.73	0.39	9.95	0.50	10.61	0.63	10.90
OE-PLR-MSE	0.58	5.83	0.67	7.22	0.80	8.26	0.86	9.74
EE-MSE	0.29	11.73	0.13	13.43	0.12	14.34	0.03	15.73
Valero (offline)	0.97	2.28	1.22	2.04	1.91	1.65	2.02	1.44
Favrot (modified): $\kappa = 12$	0.77	5.36	0.72	4.91	0.92	4.75	1.28	4.37
Favrot (modified): $\kappa = 16$	0.62	6.35	0.61	5.38	0.76	5.14	1.12	4.43

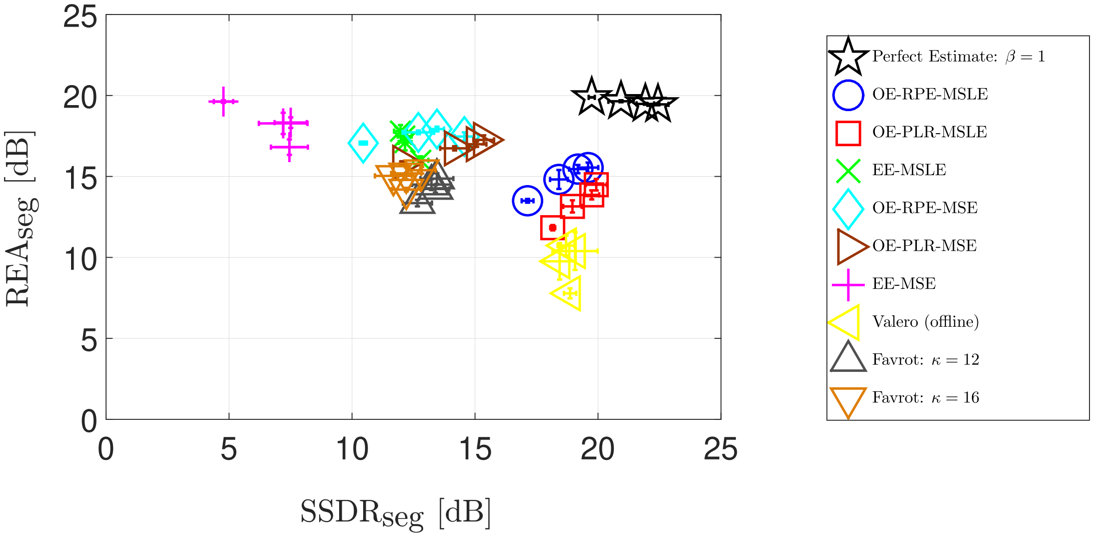

Fig. 12. Plot of segmental REA vs segmental SSDR obtained for RIRs measured in four different rooms for all considered parameter estimation methods.

not completely cancelled, i.e. a small amount of early residual echo remains due to filter misalignment ($R_E \neq 0$). In addition, near-end speech and background noise are present, with the near-end signal-to-noise ratio set to 40 dB. In order to obtain a fair comparison of the segmental performance metrics for all measured RIRs, all RIRs have been scaled appropriately such that the speech-to-residual echo ratio (SRER) is equal to 10 dB. The reverberation parameters $\underline{\theta}(k)$ are estimated only during periods of near-end speech absence, i.e. during the first 25 secs, and when the AEC error PSD Φ_e is at least 3 dB above the background noise PSD Φ_v, as during these periods the AEC error PSD Φ_e is predominantly composed of the LRE PSD Φ_{r_L}. As Φ_{r_L} is not directly observable in practice, it is approximated in (30) by Φ_e during these periods.

For this realistic setting, we compare the LSD, REA$_{seg}$ and SSDR$_{seg}$ scores obtained using the OE-RPE, OE-PLR and EE methods (for both the MSE and MSLE cost functions) with the state-of-the-art methods in [16] (offline version) and [17] (modified online version). Additionally, we also compare the estimated reverberation time $\hat{T}_{60}$ with the (true) T_{60} obtained by line-fitting.

Fig. 11 shows the LSD scores obtained using all considered methods for the measured RIRs in each room. Each point in this figure corresponds to the average LSD score obtained for all RIRs in a specific room, while the error bars depict the standard deviation across these RIRs. It can be observed that the proposed OE-RPE-MSLE and OE-PLR-MSLE methods outperform all other online parameter estimation methods, and are even slightly better than the offline method proposed in [16]. In addition, Table IV breaks down all average LSD scores into under- and over-estimation scores. Firstly, it can be observed that among all proposed estimation methods, the OE-RPE-MSLE and OE-PLR-MSLE methods yield similar under- and over-estimation scores. Although the other proposed methods and the modified Favrot method yield smaller under-estimation scores than the OE-RPE-MSLE and OE-PLR-MSLE methods, they yield considerably larger over-estimation scores. Finally, for the offline method proposed in [16], both the under- and over-estimation scores are slightly larger than for the online OE-RPE-MSLE and OE-PLR-MSLE methods (except for under-estimation scores in the lab and over-estimation scores in the echoic room).

Fig. 12 shows the REA$_{seg}$ scores against the SSDR$_{seg}$ scores obtained using all considered methods. Each point in this figure corresponds to the average result obtained for all RIRs in a specific room, while the error bars on the x and y-axes depict the standard deviations across these RIRs for the SSDR$_{seg}$ and REA$_{seg}$ scores, respectively. For comparison, we also included the results obtained using the perfect LRE PSD estimate $\hat{\Phi}_{r_L} = \Phi_{r_L}$ and an over-estimation factor $\beta = 1$, which yields the best possible performance in terms of maximizing both the REA$_{seg}$

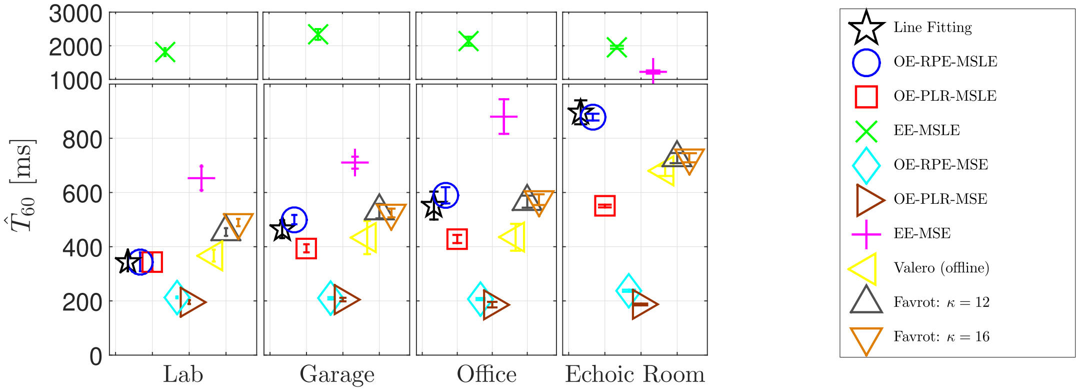

Fig. 13. Plot of $\hat{T}_{60}$ vs T_{60} (line-fitting) for RIRs measured in four different rooms for all considered parameter estimation methods.

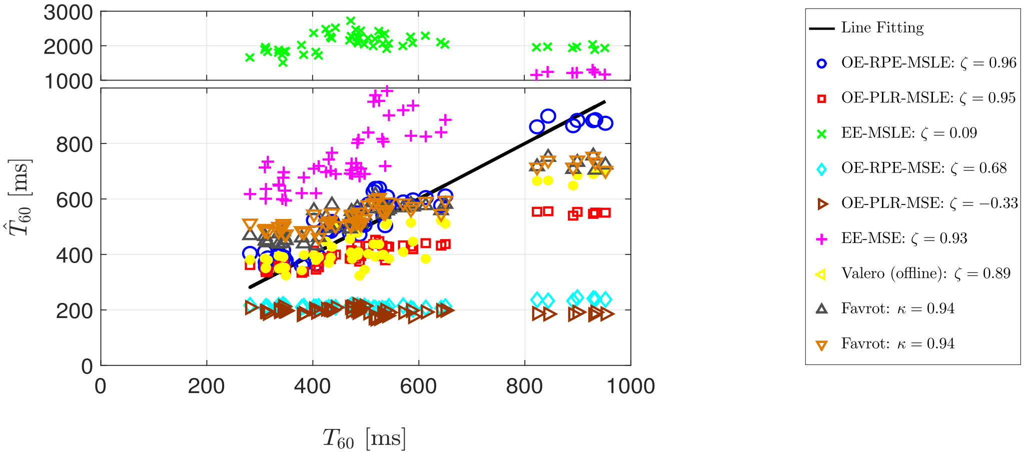

Fig. 14. Correlation between $\hat{T}_{60}$ obtained using each considered parameter estimation method and T_{60} (line-fitting) for all measured RIRs.

and SSDR_{seg} scores. As expected, it can be observed that a large LSD over-estimation score (see Table IV) leads to large residual echo attenuation at the expense of large near-end speech distortion, while an opposite effect can be observed for a large LSD under-estimation score. The proposed online OE-RPE-MSLE and OE-PLR-MSLE methods as well as Valero's offline method yield significantly better SSDR_{seg} scores as compared to the other methods (about 5–10 dB), while not losing too much in terms of the REA_{seg} score (about 2–3 dB). Overall, the OE-RPE-MSLE method yields the best performance amongst all considered parameter estimation methods, i.e. both its REA_{seg} as well as its SSDR_{seg} score are closest to the scores obtained for the perfect LRE PSD estimate.

Fig. 13 shows the estimated reverberation time $\hat{T}_{60}$ obtained using all considered methods for the measured RIRs in each room. Each point in this figure corresponds to the average result obtained for all RIRs in a specific room, while the error bars depict the standard deviation across these RIRs. For comparison, the (true) T_{60} values obtained by line-fitting on the measured RIRs have also been included. It can be clearly observed that the OE-RPE-MSLE method yields the most accurate and consistent T_{60} estimate across all rooms. On the one hand, the OE-PLR-MSLE method, Valero's method and Favrot's method perform rather similarly, i.e. slightly over-estimating the T_{60} for the lower range (250–500 ms) but under-estimating the T_{60} for the higher range (600–900 ms). On the other hand, the EE method for both cost functions fails completely and significantly over-estimates the T_{60}, while the OE-RPE-MSE and OE-PLR-MSE methods significantly underestimate the T_{60}. Additionally, in Fig. 14 we plot the estimated and true T_{60} values for all 55 measured RIRs and compute the correlation coefficient ζ between these values for each considered method. It can be seen that the proposed OE-RPE-MSLE method yields the largest correlation coefficient ($\zeta = 0.96$), followed by the proposed OE-PLR-MSLE method ($\zeta = 0.95$) and Favrot's method ($\zeta = 0.94$).

In conclusion, based on the results obtained for this realistic setting, the proposed OE-RPE-MSLE method outperforms all other considered (online and offline) parameter estimation methods in terms of LRE PSD and T_{60} estimation accuracy, while yielding the largest SSDR_{seg} score and hardly compromising on the REA_{seg} score compared to the perfect LRE PSD estimate.

VII. Conclusion

In this paper, we considered late residual echo suppression by jointly estimating the parameters of an exponentially decaying reverberation model using online signal-based methods. The OE and EE methods, which were originally proposed to estimate the coefficients of time-domain IIR filters, were used on PSDs to jointly estimate the reverberation scaling and decay parameters by minimizing either the MSE or the MSLE cost function. For both methods, gradient-descent-based algorithms were derived to simultaneously update both parameters during periods of near-end speech absence. The estimated parameters were then used in a recursive filter structure to generate the corresponding LRE PSD estimate. The different methods (OE/EE), cost functions (MSE/MSLE) and gradient-descent-based algorithms (RPE/PLR) were compared with state-of-the-art signal-based methods, both in an idealistic as well as in a realistic setting. For both considered settings, the proposed OE-RPE-MSLE and OE-PLR-MSLE methods consistently outperformed all other considered methods in terms of LRE PSD estimation accuracy. Moreover, across all considered scenarios the OE-RPE-MSLE method yielded the most accurate T_{60} estimates. The EE method failed to accurately estimate the LRE PSD and T_{60} across all scenarios, while both OE and EE methods for the MSE cost function failed to accurately estimate the T_{60}. For the realistic setting, the proposed OE-RPE-MSLE and OE-PLR-MSLE methods resulted in the smallest near-end speech distortion after applying the postfilter, while delivering a large residual echo suppression.

APPENDIX A
DERIVATION OF MODEL FOR LATE RESIDUAL ECHO PSD

We adopt the methodology used in [34] and [35] to derive the recursive expression for λ_{r_L} in (14), as well as expressions for the reverberation parameters A and B in terms of the RIR model parameters σ_L^2 and ρ in (15) and (16). The energy envelope of the late part of the stochastic RIR h in (12) is given as:

$$E\{h^2(i)\} = \sigma_L^2 \cdot e^{-2\rho\,(i-N)},\ N \le i < N_h, \tag{53}$$

where $E\{\cdot\}$ denotes spatial expectation, i.e. the ensemble average over different realizations of the stochastic process h. As the LRE signal r_L is given as:

$$r_L(n) = \sum_{i=N}^{N_h-1} h(i) \cdot x(n-i), \tag{54}$$

its auto-correlation at lag τ for one realization of h is defined as:

$$\begin{aligned} a_{r_L r_L}(n, n+\tau; h) &= \mathscr{E}\{r_L(n) \cdot r_L(n+\tau)\} \\ &= \sum_{i=N}^{N_h-1} \sum_{j=N}^{N_h-1} h(i) \cdot h(j) \cdot \mathscr{E}\{x(n-i) \cdot x(n-j+\tau)\} \\ &= \sum_{i=N}^{N_h-1} \sum_{j=N}^{N_h-1} h(i) \cdot h(j) \cdot a_{xx}(n-i, n-j+\tau), \end{aligned} \tag{55}$$

where $a_{xx}(n, n+\tau)$ denotes the auto-correlation of the far-end signal $x(n)$ at lag τ. Assuming that h and x are mutually independent, the spatial average of (55) over all realizations of h can be computed using (53) as:

$$\begin{aligned} a_{r_L r_L}(n, n+\tau) &= E\{a_{r_L r_L}(n, n+\tau; h)\} \\ &= \sum_{i=N}^{N_h-1} \sum_{j=N}^{N_h-1} E\{h(i) \cdot h(j)\} \cdot a_{xx}(n-i, n-j+\tau) \\ &= \sigma_L^2 \cdot e^{2\rho\,N} \cdot \sum_{i=N}^{N_h-1} e^{-2\rho\,i} \cdot a_{xx}(n-i, n-i+\tau), \end{aligned} \tag{56}$$

since $E\{h(i) \cdot h(j)\} = 0$ if $i \neq j$. Evaluating (56) at time instant $n - F$, with $F \ll N_h$, gives:

$$\begin{aligned} &a_{r_L r_L}(n-F, n-F+\tau) \\ &\quad = \sigma_L^2 \cdot e^{2\rho\,N} \cdot \sum_{i=N}^{N_h-1} e^{-2\rho\,i} \cdot a_{xx}(n-F-i, n-F-i+\tau) \\ &\quad \approx \sigma_L^2 \cdot e^{2\rho\,N} \cdot \sum_{i=N+F}^{N_h-1} e^{-2\rho\,(i-F)} \cdot a_{xx}(n-i, n-i+\tau). \end{aligned} \tag{57}$$

Using (56) and (57), the auto-correlation of the LRE signal $a_{r_L r_L}$ can be computed recursively as:

$$\begin{aligned} a_{r_L r_L}(n, n+\tau) &= e^{-2\rho\,F} \cdot a_{r_L r_L}(n-F, n-F+\tau) \\ &+ \sigma_L^2 \cdot e^{2\rho\,N} \cdot \sum_{i=N}^{N+F-1} e^{-2\rho\,i} \cdot a_{xx}(n-i, n-i+\tau). \end{aligned} \tag{58}$$

If we assume the signal x to be stationary over F samples, with F the STFT frameshift, (58) can be rewritten as:

$$\begin{aligned} a_{r_L r_L}(n, n+\tau) &= e^{-2\rho\,F} \cdot a_{r_L r_L}(n-F, n-F+\tau) \\ &+ \sigma_L^2 \cdot \left(\frac{1-e^{-2\rho\,F}}{1-e^{-2\rho}}\right) \cdot a_{xx}(n-N, n-N+\tau). \end{aligned} \tag{59}$$

Using the Wiener-Khinchin theorem, (59) can be expressed in terms of true PSDs as:

$$\lambda_{r_L}(k, \ell) = A \cdot \lambda_x(k, \ell - G) + B \cdot \lambda_{r_L}(k, \ell - 1), \tag{60}$$

where $G = \lfloor \frac{N}{F} \rfloor$ and the parameters A and B are equal to:

$$A = \sigma_L^2 \cdot \left(\frac{1-e^{-2\rho\,F}}{1-e^{-2\rho}}\right), \tag{61}$$

$$B = e^{-2\rho F}. \tag{62}$$

APPENDIX B
MODIFIED VERSION OF PSD ESTIMATION METHOD IN [17]

We denote the parameters estimated using the modified version of Favrot's method [17] as $\hat{\underline{\theta}}^{\mathrm{F}}$. The parameter A^{F} corresponds to the initial power of the residual echo and is estimated as:

$$\hat{A}_N^{\mathrm{F}}(k, \ell) = \frac{\mathscr{E}\{\tilde{\Phi}_e(k, \ell) \cdot \tilde{\Phi}_{x_N}(k, \ell)\}}{\mathscr{E}\{\tilde{\Phi}_{x_N}(k, \ell) \cdot \tilde{\Phi}_{x_N}(k, \ell)\}}, \tag{63}$$

where $\tilde{\Phi}_{x_N}(k, \ell) = |X_N(k, \ell)|^2 - \Phi_{x_N}(k, \ell)$ and $\tilde{\Phi}_e(k, \ell) = |E(k, \ell)|^2 - \Phi_e(k, \ell)$ represent the temporal fluctuations of the

PSD of the N-sample delayed far-end signal $x_N(n) = x(n - N)$ and the AEC error signal $e(n)$, respectively. The far-end signal is delayed so as to temporally align it with the LRE component in the AEC error signal. Thus, the numerator in (63) is the cross-correlation between the temporal fluctuations of the delayed far-end signal PSD and the AEC error PSD, while the denominator is the auto-correlation of the temporal fluctuations of the delayed far-end signal PSD. The decay rate is estimated by computing (63) for two different delays M and $M + P$, where M should be chosen such that $\hat{A}^{\mathrm{F}}_{M}$ can be associated with the late reverberant part of the RIR h and P corresponds to a delay of κ frames, i.e.:

$$\hat{B}^{\mathrm{F}}(k,\ell) = \left(\frac{\hat{A}^{\mathrm{F}}_{M+P}(k,\ell)}{\hat{A}^{\mathrm{F}}_{M}(k,\ell)}\right)^{1/\kappa}. \tag{64}$$

REFERENCES

[1] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. New York, NY, USA: Wiley, 2004.

[2] S. Haykin, *Adaptive Filter Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.

[3] C. Breining *et al.* "Acoustic echo control - An application of very-high-order adaptive filters," *IEEE Signal Process. Mag.*, vol. 16, no. 4, pp. 42–69, Jul. 1999.

[4] H. Huang, C. Hofmann, W. Kellermann, J. Chen, and J. Benesty, "A multiframe parametric Wiener filter for acoustic echo suppression," in *Proc. IEEE Int. Workshop Acoust. Signal Enhanc.*, Xi'an, China, 2016, pp. 1–5.

[5] C. Beaugeant, V. Turbin, P. Scalart, and A. Gillore, "New optimal filtering approaches for hands-free telecommunication terminals," *Signal Process.*, vol. 64, no. 1, pp. 33–47, 1998.

[6] S. Gustafsson, R. Martin, P. Jax, and P. Vary, "A psychoacoustic approach to combined acoustic echo cancellation and noise reduction," *IEEE Trans. Speech Process.*, vol. 10, no. 5, pp. 245–256, Jul. 2002.

[7] V. Turbin, A. Gilloire, and P. Scalart, "Comparison of three postfiltering algorithms for residual acoustic echo reduction," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Munich, Germany, 1997, pp. 307–310.

[8] C. Lee, J. Shin, and N. Kim, "DNN-based residual echo suppression," in *Proc. 16th Annu. Conf. Int. Speech Commun. Assoc.*, Dresden, Germany, 2015, pp. 1775–1779.

[9] I. Schalk-Schupp, F. Faubel, M. Buck, and A. Wendemuth, "Approximation of a nonlinear distortion function for combined linear and nonlinear residual echo suppression," in *Proc. IEEE Int. Workshop Acoust. Signal Enhanc.*, Xi'an, China, 2016, pp. 1–5.

[10] J. Franzen and T. Fingscheidt, "An efficient residual echo suppression for multi-channel acoustic echo cancellation based on the frequency-domain adaptive Kalman filter," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Calgary, Canada, 2018, pp. 226–230.

[11] G. Carbajal, R. Serizel, E. Vincent, and E. Humbert, "Multiple-input neural network-based residual echo suppression," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Calgary, Canada, 2018, pp. 231–235.

[12] E. Habets, I. Cohen, S. Gannot, and P. Sommen, "Joint dereverberation and residual echo suppression of speech signals in noisy environments," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1433–1451, Nov. 2008.

[13] J. Polack, *La Transmission de l'énergie Sonore Dans Les Salles*. Mans, France, Université du Maine, 1988.

[14] E. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Process. Lett.*, vol. 16, no. 9, pp. 770–774, Sep. 2009.

[15] G. Enzner, "A Model-based optimum filtering approach to acoustic echo control: Theory and practice," Ph.D. dissertation, IND, RWTH Aachen University, Aachen, Germany, 2006.

[16] M. Valero, E. Mabande, and E. Habets, "Signal-based late residual echo spectral variance estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Florence, Italy, 2014, pp. 5914–5918.

[17] A. Favrot, C. Faller, and F. Küch, "Modeling late reverberation in acoustic echo suppression," in *Proc. Int. Workshop Acoust. Signal Enhanc.*, Aachen, Germany, 2012, pp. 1–4.

[18] N. Desiraju, S. Doclo, M. Buck, T. Gerkmann, and T. Wolff, "On determining optimal reverberation parameters for late residual echo suppression," in *Proc. AES Conf. Dereverber. Reverber. Audio, Music, Speech*, Leuven, Belgium, 2016, pp. 1–8.

[19] J. Shynk, "Adaptive IIR filtering," *IEEE ASSP Mag.*, vol. 6, no. 2, pp. 4–21, Apr. 1989.

[20] Y. Tomita, A. Damen, and P. Van Den Hof, "Equation error versus output error methods," *Ergonomics*, vol. 35, nos. 5/6, pp. 551–564, 1992.

[21] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, Jul. 2001.

[22] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 466–475, Sep. 2003.

[23] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 4, pp. 1383–1393, May 2012.

[24] H. Kuttruff, *Room Acoustics*. London, U.K.: Spon Press, 2000.

[25] H. Schepker and S. Doclo, "Least-squares estimation of the common pole-zero filter of acoustic feedback paths in hearing aids," *IEEE/ACM Trans. Audio, Speech Lang. Process.*, vol. 24, no. 8, pp. 1334–1347, Aug. 2016.

[26] H. Schepker and S. Doclo, "A semidefinite programming approach to min-max estimation of the common part of acoustic feedback paths in hearing aids," *IEEE/ACM Trans. Audio, Speech Lang. Process.*, vol. 24, no. 2, pp. 366–377, Feb. 2016.

[27] T. Soderstrom and P. Stoica, "Some properties of the output error method," *Automatica*, vol. 18, no. 1, pp. 93–99, 1982.

[28] M. Nayeri, "A weaker sufficient condition for the unimodality of error surfaces associated with exactly matching adaptive IIR filters," *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, California, 1988, pp. 35–38.

[29] S. Stearns, "Error surfaces of recursive adaptive filters," *IEEE Trans. Circuits Syst.*, vol. CAS-28, no. 6, pp. 603–606, Jun. 1981.

[30] T. Soderstrom, "On the uniqueness of maximum likelihood identification," *Automatica*, vol. 11, no. 2, pp. 193–197, 1975.

[31] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, and N. Dahlgren, "DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM," *Nat. Inst. Standards Technol.*, vol. 93, p. 27403, 1990.

[32] M. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 37, pp. 409–412, 1965.

[33] T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model," *EURASIP J. Advances Signal Process.*, vol. 2005, no. 7, pp. 1110–1126, 2005.

[34] K. Lebart and J. Boucher, "A new method based on spectral subtraction for speech dereverberation," *Acta Acoust.*, vol. 87, pp. 359–366, 2001.

[35] E. Habets, "Speech dereverberation using statistical reverberation models," *Speech Dereverber.* New York, NY, USA: Springer, 2010, pp. 57–93.

Naveen Kumar Desiraju received the B.Tech. degree in electrical engineering from the Indian Institute of Technology Roorkee, Roorkee, India, in 2011, and the M.Sc. degree in signal processing and machine intelligence from the University of Surrey, Guildford, U.K., in 2012. He is currently working toward the Ph.D. degree at the Signal Processing Group, Department of Medical Physics and Acoustics, Carl von Ossietzky University, Oldenburg, Germany. From 2011 to 2012, he was a Commonwealth scholar with the University of Surrey, Guildford, U.K., where he won the Farzin Mokhtarian prize for best M.Sc. student in the Department of Electrical and Electronic Engineering. From 2013 to 2017, he was a Doctoral Researcher with the Acoustic Speech Enhancement team, Nuance Communications Deutschland GmbH, Ulm, Germany, on a Marie Skłodowska- Curie fellowship as part of the European Commission's DREAMS project (Dereverberation and REverberation of Audio Music and Speech). He is an Associate Engineer in Automatic Speech Recognition with Harman Connected Services GmbH, Garching bei München, Germany. His research interests are in multi-channel speech enhancement, adaptive filtering, and machine learning.

Simon Doclo (S'95–M'03–SM'13) received the M.Sc. degree in electrical engineering and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1997 and 2003, respectively. From 2003 to 2007, he was a Postdoctoral Fellow with the Research Foundation Flanders, the Department of Electrical Engineering (Katholieke Universiteit Leuven), and the Cognitive Systems Laboratory (McMaster University, Canada). From 2007 to 2009, he was a Principal Scientist with NXP Semiconductors in Leuven, Belgium. Since 2009, he has been a full Professor with the University of Oldenburg, Germany, and Scientific Advisor for the Branch Hearing, Speech and Audio Technology of the Fraunhofer Institute for Digital Media Technology. His research interests include signal processing for acoustical and biomedical applications, more specifically microphone array processing, speech enhancement, active noise control, acoustic sensor networks, and hearing aid processing.

Prof. Doclo is a recipient of several best paper awards (International Workshop on Acoustic Echo and Noise Control 2001, EURASIP Signal Processing 2003, IEEE Signal Processing Society 2008, VDE Information Technology Society 2019). He is member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing, the EURASIP Technical Area Committee on Acoustic, Speech and Music Signal Processing, and the EAA Technical Committee on Audio Signal Processing. He was and is involved in several large-scale national and European research projects (ITN DREAMS, Cluster of Excellence Hearing4all, CRC Hearing Acoustics). He was Technical Program Chair of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics in 2013 and Chair of the ITG Conference on Speech Communication in 2018. In addition, he served as a Guest Editor for several special issues (IEEE SIGNAL PROCESSING MAGAZINE, Elsevier Signal Processing) and is Associate Editor for the IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING and EURASIP Journal on Advances in Signal Processing.

Markus Buck (M'13) received the Dipl.-Ing. degree in electrical engineering and the Ph.D. degree from Ulm University, Ulm, Germany, in 1998 and 2004, respectively.

From 1998 to 2009, he was building up expertise in the domain of speech enhancement at Temic Speech Dialog Systems and Harman Becker Automotive Systems, Ulm, Germany. Since 2009, he has been a Research Manager with the Nuance Communications Deutschland GmbH and Cerence Inc., Ulm, Germany, leading the technology development in acoustic speech enhancement for hands-free telephony, speech recognition, and in-car communication. His main research interests include multi-channel signal processing, adaptive filtering, and neural network based approaches for speech signal processing.

Tobias Wolff received the Dipl.-Ing. degree and the Dr.-Ing. degree in electrical engineering and communications from the Signal Processing Group, Technische Universität Darmstadt, Darmstadt, Germany, in 2006 and 2011, respectively. In 2005 and 2007, he was a Visiting Researcher with the Image Processing Laboratory, University of California Santa Barbara, USA, working on subjective perception of video coding artifacts. In April 2009, he joined the Department of Speech Signal Enhancement, Nuance Communications Deutschland GmbH, Ulm, Germany. Since 2017, he has been a Principal Researcher with Nuance in the area of multi-microphone acoustic speech enhancement. Since 2019, he has been with Cerence Inc. in the same domain. His main scientific research interests include beamforming, acoustic source localization and dereverberation of speech signals for robust speech recognition in the home environment. He was a Supervisor in the European research project DREAMS.