On Determining Optimal Reverberation Parameters for Late Residual Echo Suppression

Naveen Kumar Desiraju[1], Simon Doclo[2], Markus Buck[1], Timo Gerkmann[2], Tobias Wolff[1]

[1]*Acoustic Speech Enhancement Research, Nuance Communications Deutschland GmbH, Ulm, Germany*

[2]*University of Oldenburg, Dept. of Medical Physics and Acoustics and Cluster of Excellence Hearing4All, Oldenburg, Germany*

Correspondence should be addressed to Naveen Kumar Desiraju (naveen.desiraju@nuance.com)

ABSTRACT

When deploying acoustic echo cancellation systems in large rooms, using short filters may result in significant amount of residual echo caused by room reverberation. In this paper, we model the late residual echo as exponentially decaying and use a parametric IIR filter to estimate its power in the subband domain for application in residual echo suppression. Working in an offline system identification setup, the problem of finding the optimal parameters of the IIR filter is addressed, with an analysis conducted on the performance of two parameter estimation methods: output error and equation error. The late residual echo power estimates obtained using the two methods are furthermore judged using the mean squared error and the mean squared log error cost functions. Results indicate that minimizing the mean squared log error for the output error method provides accurate estimates for the late residual echo power and the reverberation decay parameter.

1. INTRODUCTION

In acoustic echo cancellation (AEC) applications, it is often desirable to use short AEC filters for reasons of computational complexity. For a room with a room impulse response (RIR) shorter than the length of the AEC filter, the entirety of the residual echo obtained may be attributed to filter mismatch [1]. However, for a room with a long RIR and/or large reverberation time (T_{60}), a significant amount of late residual echo remains due to the insufficient length of the AEC filter. This late residual echo signal is typically suppressed using a residual echo suppression (RES) filter, which relies on an accurate estimate of the late residual echo power spectral density (PSD) [1]. The RES filter takes on increased importance in handling the acoustic echo effectively when short AEC filters are employed.

Different methods have been proposed for estimating the late residual echo PSD. In [2], the uncompensated part of the RIR is modeled as exponentially decaying using a statistical reverberation model based on the reverberation scaling and decay parameters. The parameters are estimated using the coefficients of the AEC filter (channel-based approach), which is assumed to be converged, and then used to estimate the late residual echo PSD. In [3], a pure echo suppression system is considered and the late residual echo PSD is estimated via recursive smoothing using the decay parameter, which is estimated using a signal-based approach. In [4], the late part of the RIR is modeled similarly as in [2], with the parameters estimated using a signal-based approach by minimizing different cost functions. The decay parameter is estimated by minimizing the mean squared log error (MSLE), with this estimate used in the estimation of the scaling parameter by minimizing the mean squared error (MSE).

As in [2], in this paper we model the uncompensated part of the RIR as exponentially decaying, and use a non-adaptive IIR filter based on the reverberation scaling and decay parameters to estimate the late residual echo PSD. The parameters of this IIR filter are estimated jointly in offline batch mode using two signal-based methods: the output error method and the equation error method. The output error method is frequently used for identifying linear systems and, for adaptive IIR filters, often displays local minima in the cost function on account of being non-linear in the parameters [5, 6, 7]. As an alternative, the equation error method is often employed since

the minimum of its convex error surface can be easily found [5, 6, 7]. In this paper, we look at these parameter estimation methods for a non-adaptive IIR filter only. For each method, we minimize both the MSE and MSLE cost functions and analyze the parameters thus obtained. The objective of this paper is to investigate if the equation error method can estimate the same reverberation parameters as the output error method in offline batch mode for our late residual echo PSD estimation setup in the subband domain.

The paper is organized as follows; in Section 2, our experimental setup is presented. The proposed model for estimating the late residual echo PSD using reverberation parameters is presented in subsection 2.1, while the MSE and MSLE cost functions are presented in subsection 2.2. In Section 3, we discuss the output error and equation error methods for estimating the reverberation parameters. In Section 4, we present the simulation setup followed by a discussion of the results in Section 5. Conclusions are presented in Section 6.

2. EXPERIMENTAL SETUP

We consider a loudspeaker-enclosure-microphone (LEM) system with a time-invariant RIR h in which the reference signal x is played out. In the absence of local speech and noise signals, the microphone signal y can be given at discrete-time sample n as:

$$y(n) = \sum_{i=0}^{\infty} h(i) \cdot x(n-i), \tag{1}$$

where i denotes the filter-sample index. The residual echo signal r can be given at sample n as:

$$\begin{aligned} r(n) &= \sum_{i=0}^{\infty} \{h(i) - \hat{h}(i)\} \cdot x(n-i), \\ &= \sum_{i=0}^{N-1} \Delta h(i) \cdot x(n-i) + \sum_{i=N}^{\infty} \Delta h(i) \cdot x(n-i), \\ &= r_E(n) + r_L(n), \end{aligned} \tag{2}$$

where $\hat{h}$ denotes the time-invariant AEC filter and Δh denotes the difference system. Here, N is chosen so as to cover just the direct sound component and the early reflections in h. The residual echo r can then be seen as being composed of the early residual echo r_E and the late residual echo r_L.

In this paper, we assume

$$\begin{aligned} \hat{h}(i) &= h(i) \quad \forall\, i < N \quad \text{and} \\ \hat{h}(i) &= 0 \quad\;\; \forall\, i \geq N, \end{aligned}$$

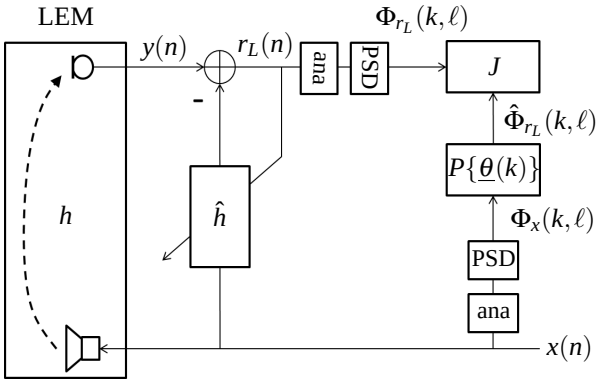

Figure 1: Setup for late residual echo PSD estimation.

i.e. no filter misalignment for the early part ($r_E(n) = 0$) and r_L existing solely due to using a short AEC filter. Thus the early part of the Δh-system is modeled using zeros, while we assume Polack's model [8] for the late part:

$$\Delta h(i) = \begin{cases} 0, & 0 \leq i < N \\ w_L(i) \cdot e^{-\rho_b (i-N)}, & i \geq N. \end{cases} \tag{3}$$

The late part is modeled as exponentially decaying *reverberation*, with w_L denoting a zero-mean white Gaussian noise (WGN) process (with variance σ_L^2) and ρ_b denoting the broadband decay rate, which is related to the T_{60} as:

$$\rho_b = \frac{3 \cdot \ln 10}{f_s \cdot T_{60}}, \tag{4}$$

where f_s is the sampling rate. The setup is shown in Figure 1, where the residual echo signal contains just the late residual echo component.

2.1. Parametric Model

We use a recursive expression, similar to the one shown in [2, 4], for the late residual echo PSD λ_{r_L} in subband k and frame ℓ using the broadband reverberation scaling parameter A_b and the broadband reverberation decay parameter B_b:

$$\lambda_{r_L}(k,\ell) = A_b \cdot \lambda_x(k, \ell - G) + B_b \cdot \lambda_{r_L}(k, \ell - 1). \tag{5}$$

Here, λ_x denotes the reference PSD and $G = \lfloor \frac{N}{F} \rfloor$ corresponds to the length of the AEC filter in frames, with F denoting the frameshift (in samples) used in the STFT operation. The broadband reverberation parameters are related to the parameters σ_L^2 and ρ_b of the Δh-system in

Equation 3, and have been derived in a manner similar to the one shown in [9]:

$$A_b = \sigma_L^2 \cdot \left(\frac{1 - e^{-2 \cdot \rho_b \cdot F}}{1 - e^{-2 \cdot \rho_b}} \right), \quad (6)$$

$$B_b = e^{-2 \cdot \rho_b \cdot F}. \quad (7)$$

The derivation has been omitted here for the sake of brevity.

Based on Equation 5, we choose to treat the task of estimating the late residual echo PSD as a system identification problem. In practice, the PSDs λ_x and λ_{r_L} are approximated by computing the squared magnitudes of the STFT of the signals x and r_L, and are denoted by Φ_x and Φ_{r_L} respectively. The quantities Φ_x and Φ_{r_L} do not perfectly estimate the PSDs λ_x and λ_{r_L}, and hence suffer from estimation noise.

Referring to Figure 1, Φ_x is filtered by the parametric IIR filter $P\{\underline{\theta}(k)\}$ to generate the estimate $\hat{\Phi}_{r_L}$, where $\underline{\theta}(k)$ represents the subband parameter set:

$$\underline{\theta}(k) = \begin{bmatrix} A(k) \\ B(k) \end{bmatrix}. \quad (8)$$

Figure 2 depicts the IIR filter $P\{\underline{\theta}(k)\}$, with z^{-1} denoting a 1-frame delay. The direct branch of the filter involves delaying $\Phi_x(k,\ell)$ by G frames and scaling it by $A(k)$, while the recursive branch contains a first-order recursion using $B(k)$.

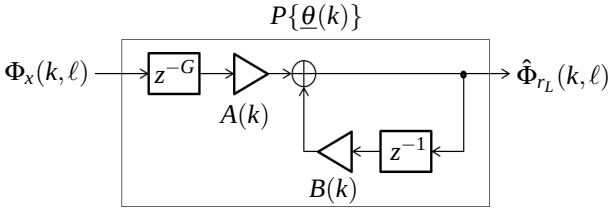

Figure 2: IIR filter for generating the late residual echo PSD estimate from the reference PSD.

2.2. Cost Functions

We evaluate the estimate $\hat{\Phi}_{r_L}(k,\ell)$ by computing the following cost functions:

- The well-known MSE (mean squared error) cost function is computed for each subband as the squared difference between the target and the estimate, averaged over the batch size of L frames:

$$J_{MSE}(k) = \frac{1}{L} \sum_{\ell=0}^{L-1} \left[\Phi_{r_L}(k,\ell) - \hat{\Phi}_{r_L}(k,\ell) \right]^2. \quad (9)$$

- The MSLE (mean squared log error) cost function is computed for each subband as the squared difference between the natural logarithms of the target and the estimate, averaged over L frames:

$$\begin{aligned} J_{MSLE}(k) &= \frac{1}{L} \sum_{\ell=0}^{L-1} \left[\ln \Phi_{r_L}(k,\ell) - \ln \hat{\Phi}_{r_L}(k,\ell) \right]^2 \\ &= \frac{1}{L} \sum_{\ell=0}^{L-1} \left[\ln \left(\frac{\Phi_{r_L}(k,\ell)}{\hat{\Phi}_{r_L}(k,\ell)} \right) \right]^2. \end{aligned} \quad (10)$$

3. PARAMETER ESTIMATION METHODS

We jointly estimate the reverberation parameters per subband in offline batch mode using two methods: output error and equation error. The estimated parameters are then fed into the IIR filter $P\{\underline{\theta}(k)\}$ to obtain an estimate for the late residual echo PSD.

3.1. Output Error Method

The output error (OE) method is a well-known technique used for the identification of linear systems in a variety of applications, both in the time-domain as well as in the subband-domain. Using the OE method for estimating the reverberation parameters per subband (Figure 3) is characterized by the following recursive difference equation (the superscript O denotes the OE method):

$$\hat{\Phi}^O_{r_L}(k,\ell) = \hat{A}(k) \cdot \Phi_x(k,\ell - G) + \hat{B}(k) \cdot \hat{\Phi}^O_{r_L}(k,\ell - 1). \quad (11)$$

We then evaluate the estimate $\hat{\Phi}^O_{r_L}$ using a chosen cost function J to judge the quality of the reverberation parameter estimates.

3.2. Equation Error Method

As illustrated in Figure 4, the EE method differs from the OE method by using the delayed target $\Phi_{r_L}(k,\ell - 1)$ instead of the delayed estimate $\hat{\Phi}_{r_L}(k,\ell - 1)$, thereby breaking the recursive structure and resulting in an FIR structure:

$$\hat{\Phi}^E_{r_L}(k,\ell) = \hat{A}(k) \cdot \Phi_x(k,\ell - G) + \hat{B}(k) \cdot \Phi_{r_L}(k,\ell - 1). \quad (12)$$

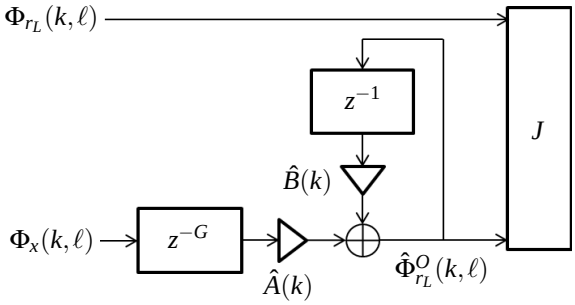

Figure 3: Reverberation parameter estimation using the output error method.

As shown in [7], under certain conditions the EE method identifies the system parameters correctly (i.e. the same parameter set as OE), which makes it particularly attractive for use in practical applications. However, as both the input Φ_x and the target Φ_{r_L} suffer from estimation noise, this may negatively affect the quality of the parameter estimates obtained.

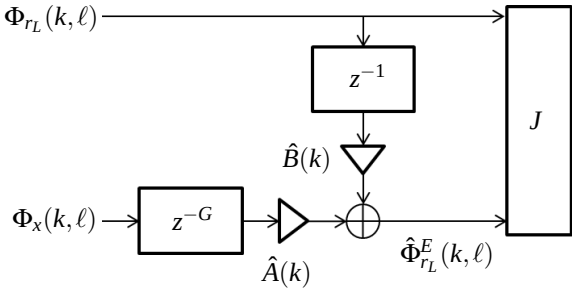

Figure 4: Reverberation parameter estimation using the equation error method.

It is important to note that the estimate $\hat{\Phi}^E_{r_L}$ cannot actually be used for RES as it is computed using the target Φ_{r_L}, while the parameters estimated using the EE method are fed into $P\{\underline{\theta}(k)\}$ to generate the late residual echo PSD estimate.

4. SIMULATION SETUP

For our simulations, we use a 1 min long clean speech signal (noiseless and dry) as the reference signal x at a sampling rate of 16 kHz. We generate RIRs exactly as described in Equation 3 for the Δh-system, with no filter mismatch and an exponentially decaying WGN tail, and convolve it with x to obtain the late residual echo signal r_L. For most RIRs, we use $10\cdot\log_{10}\left(\sigma_L^2\right) = -26$ dB and T_{60} ranging from 100 ms to 1000 ms in steps of 100 ms. Other RIRs are generated for $T_{60} = 500$ ms for σ_L^2 values of -40, -32 and-22.5 dB. Both x and r_L are converted into the STFT domain using an analysis filterbank of DFT-order $N_{\mathrm{FFT}} = 512$ and frameshift $F = 128$.

As we are looking for the optimal parameter set $\underline{\theta}_{opt}(k)$ which minimizes a given cost function J for our filter $P\{\underline{\theta}(k)\}$, we proceed to find it by conducting a grid search. We choose 100 discrete values for A and B each and through their combination generate 10^4 distinct parameter sets. The values for A are chosen from -20 to 0 dB with a resolution of 0.2 dB, while those for B are chosen using Equation 7, with T_{60} from 60-1050 ms in steps of 10 ms. For the OE method, the input Φ_x is processed in a batch (per subband) using each parameter set to generate $\hat{\Phi}^O_{r_L}$ using Equation 11, which is evaluated using J. An error surface with 10^4 grid-points is then plotted, with the parameter set $\underline{\hat{\theta}}^O_{min}(k)$ obtained by searching for the global minimum in the error surface in the k-th subband. The same operation is performed for the EE method by processing Φ_x and Φ_{r_L} in batches using Equation 12 to obtain $\underline{\hat{\theta}}^E_{min}(k)$.

5. RESULTS AND DISCUSSIONS

We present and discuss the results obtained, with error surfaces and estimated reverberation parameters presented in subsection 5.1, estimates for the parameters σ_L^2 and T_{60} presented in subsection 5.2 and the late residual echo PSD estimates presented in subsection 5.3.

5.1. Error Surfaces

Figure 5 shows the contour plots of the error surfaces obtained for the MSE (left) and the MSLE (right) cost functions for $P\{\underline{\theta}(k)\}$. The plots show the contours as a function of A and B (in dB) in subband 25 ($\approx$ 800 Hz), with the target Φ_{r_L} computed for an RIR with $\sigma_L^2 = -26$ dB and $T_{60} = 500$ ms. Each plot shows 20 contour lines, equidistant on the dB scale, with the levels of some contour lines displayed. The green circles, the red triangles and the black squares in the plots represent $\underline{\hat{\theta}}^O_{min}(k)$, $\underline{\hat{\theta}}^E_{min}(k)$ and $\underline{\theta}_b$ respectively, where $\underline{\theta}_b$ is obtained using Equations 6 and 7. We observe that the EE method is unable to find the same parameter set as the OE method in this subband, i.e.

$$\underline{\hat{\theta}}^E_{min}(k) \neq \underline{\hat{\theta}}^O_{min}(k) = \underline{\theta}_{opt}(k). \tag{13}$$

A similar result is observed across all subbands. It is interesting to note that $\underline{\theta}_{opt}(k)$ is not the same for J_{MSE} and

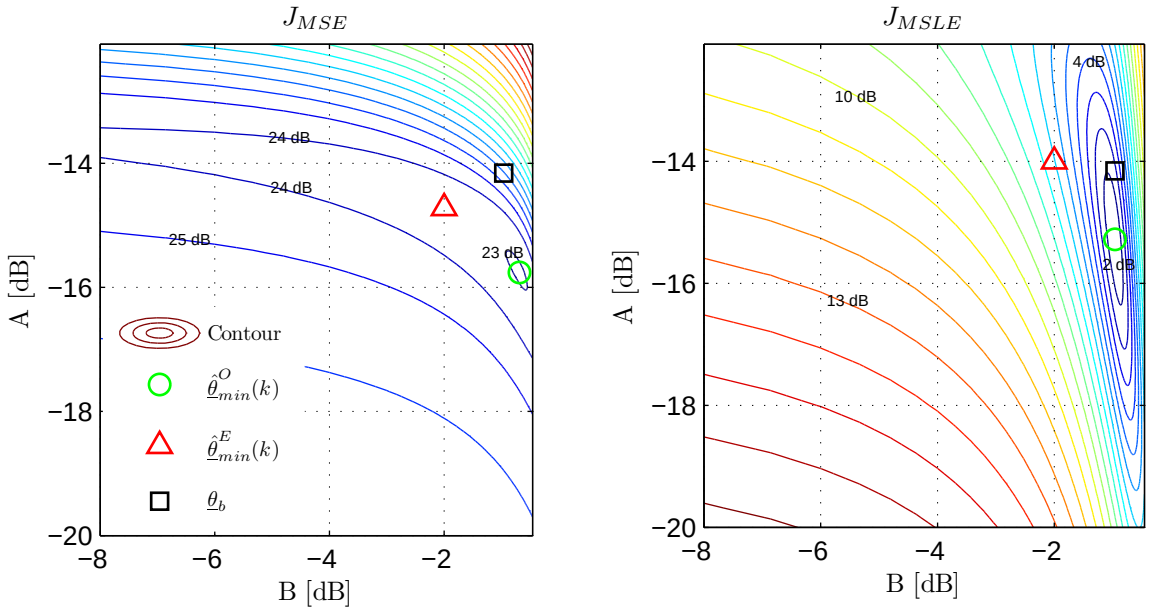

Figure 5: Contour plot of error surfaces in subband 25 for the MSE (Equation 9) and MSLE (Equation 10) cost functions obtained for the IIR filter $P\{\underline{\theta}(k)\}$ (Figure 2), for $\sigma_L^2 = -26$ dB and $T_{60} = 500$ ms. The green circles and the red triangles represent $\hat{\underline{\theta}}^{O}_{min}(k)$ and $\hat{\underline{\theta}}^{E}_{min}(k)$ respectively, i.e. the parameter sets corresponding to the global minimum in the error surfaces for the output error and the equation error methods; while the black squares represent the broadband parameter set $\underline{\theta}_b$ (from Equations 6 and 7).

J_{MSLE}, which suggests that the choice of cost function is an important factor in determining $\hat{\underline{\theta}}_{min}(k)$.

In Figure 5, we see a high density of contour lines in the vicinity of $\hat{\underline{\theta}}^{O}_{min}(k)$ for both cost functions, which indicates that the error surfaces are steep near the minimum. This indicates the presence of a unique minimum for each surface and an easily identifiable $\hat{\underline{\theta}}^{O}_{min}(k)$. Also the J_{MSLE} surface is more sensitive w. r. t. B whereas the J_{MSE} surface is more sensitive w. r. t. A. The most probable reason for this is the fact that signal frames with large power have more influence in the computation of J_{MSE} (Equation 9), thus making it more sensitive to errors in the reverberation scaling parameter, while for computing J_{MSLE}, frames with free decay (low power) are equally important as frames with high power (Equation 10), thereby making it relatively more sensitive to errors in the decay parameter.

Figure 6 shows the contour plots for the EE method used to obtain $\hat{\underline{\theta}}^{E}_{min}(k)$ for the MSE and MSLE cost functions. The contour lines in this figure are by comparison less dense in the vicinity of $\hat{\underline{\theta}}^{E}_{min}(k)$. So even though we get a unique minimum for each error surface, the $\hat{\underline{\theta}}^{E}_{min}$ corresponding to them do not stand out distinctly from the surrounding parameter sets. Also, neither cost function seems to be highly sensitive to any parameter in particular.

5.2. Parameter Estimation

In Figure 7, we plot the estimates for σ_L^2 at $T_{60} = 500$ ms, and the estimates for T_{60} at $\sigma_L^2 = -26$ dB, obtained using the OE and EE methods by minimizing both cost functions, and compare them to the broadband values. These estimates have been computed through Equations 6 and 7 by using frequency averaged $\hat{\underline{\theta}}_{min}$. The OE method gives accurate estimates for T_{60} in general, with those obtained by minimizing MSLE being more accurate than those obtained by minimizing the MSE on account of the higher sensitivity of J_{MSLE} to the decay parameter B (see Figure 5). The EE method is unable to provide accurate estimates for T_{60}, which can be explained by the lack of sensitivity of both J_{MSE} and J_{MSLE} for the EE method to the decay parameter B (see Figure 6). For σ_L^2, minimiz-

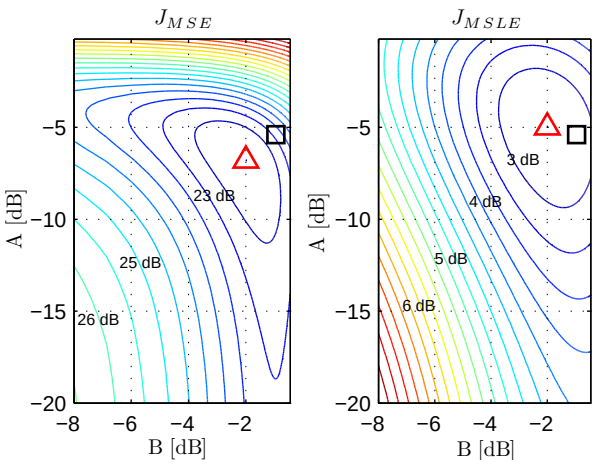

Figure 6: Contour plot of error surfaces in subband 25 for the MSE (Equation 9) and MSLE (Equation 10) cost functions obtained for the equation error method, for $\sigma_L^2 = -26$ dB and $T_{60} = 500$ ms.

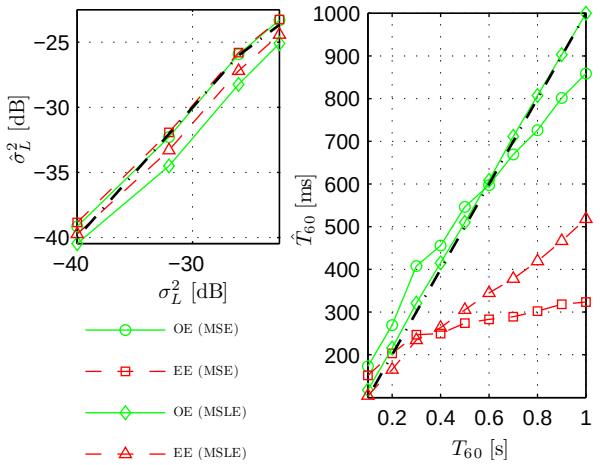

Figure 7: Comparison of estimates $\hat{T}_{60}$ and $\hat{\sigma}_L^2$ against broadband values (black dashed line), computed through Equations 6 and 7 using frequency averaged reverberation scaling and decay parameters.

ing the MSE gives accurate estimates for both OE and EE methods, while minimizing the MSLE gives a small bias in the estimates. From this figure, we can conclude that unlike the OE method, the EE method doesn't find the correct decay parameter (in the broadband sense) by minimizing either cost function, but does find the correct scaling parameter (in the broadband sense) by minimizing the MSE cost function.

The small bias obtained in $\hat{\sigma}_L^2$ by minimizing the MSLE can be explained if we assume that the real and imaginary parts of the STFT of the signals x and r_L are independent and identically normally distributed. Then, both $\Phi_x(k,\ell)$ and $\Phi_{r_L}(k,\ell)$ have $\chi^2(2)$ distributions, as they are computed using magnitude squares [10]. The estimate $\hat{\Phi}_{r_L}^E(k,\ell)$ is obtained as the sum of two $\chi^2(2)$ distributed variables (Equation 12) and so has a $\chi^2(4)$ distribution, while $\hat{\Phi}_{r_L}^O(k,\ell)$ is χ^2-distributed with even higher degrees of freedom as it is obtained through recursive smoothing (Equation 11). Thus, the ratio of $\Phi_{r_L}(k,\ell)$ and $\hat{\Phi}_{r_L}^{O/E}(k,\ell)$, used in the computation of J_{MSLE} (Equation 10), has a Fisher distribution, as it is a ratio of two χ^2 distributed variables [10]. It can be shown that the logarithm of a Fisher distributed variable with differing degrees of freedom in the numerator and denominator has a non-zero mean distribution, which may be the cause for the bias in the estimates.

5.3. Late Residual Echo PSD Estimation

After obtaining $\underline{\hat{\theta}}_{min}(k)$ using either the OE or EE method for a given cost function minimization, we feed it into the IIR filter $P\{\underline{\theta}(k)\}$ to obtain the *best-case* estimate for the late residual echo PSD:

$$\hat{\Phi}_{r_L}^{min}(k,\ell) = \hat{A}_{min}(k) \cdot \Phi_x(k,\ell) + \hat{B}_{min}(k) \cdot \hat{\Phi}_{r_L}^{min}(k,\ell-1). \quad (14)$$

To evaluate this estimate, the Log Spectral Distance (LSD) [11] between the target and the estimate is computed as follows:

$$\text{LSD} = \frac{10}{K \cdot L} \cdot \sum_{k=0}^{K-1} \sum_{\ell=0}^{L-1} |\log_{10}\left(\frac{\Phi_{r_L}(k,\ell)}{\hat{\Phi}_{r_L}^{min}(k,\ell)}\right)|, \quad (15)$$

where K denotes the total number of subbands. The absolute value operator $|\cdot|$ ensures that the errors made due to overestimation and underestimation across all subbands and frames are added up.

In Figure 8, we plot the LSD values obtained when $\hat{\Phi}_{r_L}^{min}(k,\ell)$ is generated using parameters obtained by the OE and EE methods by minimizing the MSE and MSLE cost functions. The target Φ_{r_L} used for the LSD computation is obtained for a particular T_{60} using an RIR with $\sigma_L^2 = -26$ dB. Here, we see that OE-MSLE gives the best estimates for the late residual echo PSD, closely followed by OE-MSE, while EE-MSE and EE-MSLE perform poorly. These results have a certain degree of cor-

relation with Figure 7, in that more accurate parameter estimates yield lower LSD values.

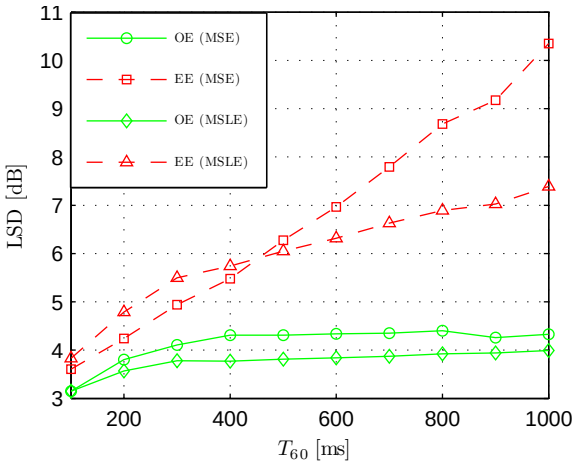

Figure 8: Comparison of Log Spectral Distance (LSD) values for $\sigma_L^2 = -26$ dB and different T_{60} values, computed using Equation 15.

6. CONCLUSIONS

In this paper, an analysis has been performed in an offline system identification setup to jointly estimate the optimal reverberation parameters (scaling and decay) in the subband domain for residual echo suppression. It has been shown that the EE method is unable to estimate the optimal parameter set for our considered model. Minimizing the MSLE cost function for the OE method gives accurate estimates for the reverberation time T_{60}, while minimizing the MSE cost function yields accurate estimates for the variance σ_L^2 of the uncompensated RIR. Based on the results obtained, we recommend that MSLE minimization should be used instead of MSE minimization for estimating the T_{60}, even when designing standalone T_{60}-estimators. The EE method is unable to estimate the T_{60} correctly with either cost function minimization. Additionally, for both the OE and EE methods, a small bias is obtained in the estimates for σ_L^2 when minimizing the MSLE cost function. Altogether, OE-MSLE minimization provides the best estimates for the reverberation parameters and the late residual echo PSD, while the EE method provides poor solutions. Based on these results, future work will focus on estimating the late residual echo PSD and the reverberation parameters using the OE method by minimizing the MSLE cost function for adaptive parametric IIR filters.

7. REFERENCES

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

[2] E.A.P. Habets, I. Cohen, S. Gannot, and P.C.W. 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.

[3] A. Favrot, C. Faller and F. Küch, "Modeling late reverberation in acoustic echo suppression", in Proc. 13th International Workshop on Acoustic Signal Enhancement, pp. 1-4, Sept. 2012.

[4] M.L. Valero, E. Mabande and E.A.P. Habets, "Signal-based Late Residual Echo Spectral Variance Estimation", in Proc. 39th IEEE International Conference on Acoustic, Speech and Signal Processing, pp. 5914-5918, May 2014.

[5] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, fourth edition, 2002.

[6] J.J. Shynk, "Adaptive IIR Filtering", IEEE ASSP Magazine, vol. 6, no. 2, pp 4-21, April 1989.

[7] Y. Tomita, A.H. Damen and P.M.J. Van Den Hof, "Equation error versus output error methods", Ergonomics, vol. 35, nos. 5/6, pp. 551-564, 1992.

[8] J. Polack, "La transmission de l'énergie sonore dans les salles", Dissertation, Université du Maine, 1988.

[9] E.A.P. Habets, "Speech dereverberation using statistical reverberation models", in *Speech Dereverberation*, P.A. Naylor and N.D. Gaubitch, Springer, London, 2010, ch. 3, pp. 57-93.

[10] L. Fahrmeir, R. Künstler, I. Pigeot and G. Tutz, *Statistik: Der Weg zur Datenanalyse*, Springer-Lehrbuch, 2004.

[11] 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.