<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Schüldt, C., Händel, P. (2014)

Decay Rate Estimators and Their Performance for Blind Reverberation Time Estimation.

IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(8): 1274-1284

<http://dx.doi.org/10.1109/TASLP.2014.2328174>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-147341>

Decay Rate Estimators and Their Performance for Blind Reverberation Time Estimation

Christian Schüldt, *Member, IEEE*, and Peter Händel, *Senior Member, IEEE*

Abstract

Several approaches for blind estimation of reverberation time have been presented in the literature and decay rate estimation is an integral part of many, if not all, of such approaches. This paper provides both an analytical and experimental comparison, in terms of the bias and variance of three common decay rate estimators; a straight-forward linear regression approach as well as two maximum-likelihood based methods. Situations with and without interfering additive noise are considered. It is shown that the linear regression based approach is unbiased if no smoothing is applied, and that the estimation variance in the absence of noise is constantly about twice that of the maximum-likelihood based methods. It is shown that the methods that do not take possible noise into account suffer from similar estimation bias in the presence of noise. Further, a hybrid method, combining the noise robustness and low computational complexity advantages of the two different maximum-likelihood based methods, is presented.

Index Terms

Blind estimation, reverberation time, decay rate, maximum-likelihood estimation.

I. INTRODUCTION

Reverberation can loosely be defined as the remaining sound in an enclosure after the source has become quiet. The reverberation characteristics are affected by the enclosure size, shape, wall reflectiveness and also possible contents such as furniture and people in a room. Generally, reverberation mitigates speech

Copyright © 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

C. Schüldt and P. Händel are with the Department of Signal Processing, ACCESS Linnaeus Centre, Royal Institute of Technology KTH, SE-100 44 Stockholm, Sweden. E-mail: {crilla,ph}@kth.se.

C. Schüldt is also with Limes Audio AB, SE-90729 Umeå, Sweden.

Manuscript received Month Day, Year; revised Month Day, Year.

intelligibility, and therefore, significant research efforts on the process of speech *de-reverberation* (i.e., the process of removing reverberation), for applications where speech is picked up by one or several microphones (such as, e.g., in hands-free telephony, just to mention one example), have been made [1]. A related method for reducing reverberation is microphone *beam-forming*, where essentially the combined filtering of multiple microphones is used to form (or increase) microphone directivity [2].

The reverberation time, T_{60} , is defined as the time required for the reverberant sound energy to decay 60 dB below that of the direct sound [3]. Knowledge of the reverberation time can be used for, e.g., estimating the volume of a room [4], speech dereverberation [5], [6] and [7], and tuning microphone beam-forming [8]. Hence, accurate T_{60} estimation is of importance in a number of applications.

A straight-forward way to estimate the reverberation time is to measure the decay curve slope of a recorded microphone signal directly after an interrupted excitation signal, which is typically in the form of broad-band or narrow-band noise [3]. Because of excitation noise fluctuations, the decay curve will differ between different trials, sometimes requiring averaging across multiple trials to obtain a reliable estimate. An alternative method, based on *backward integration* of the room impulse response (RIR), avoids multiple trials by using the relation that the ensemble average of decaying white noise in an enclosure is identical to a certain integral over the RIR. Both methods are described in the ISO 3382 standard [3].

Unfortunately, in many situations it is not possible to use a known excitation signal to obtain RIR-estimates, and the only information available is a microphone signal containing reverberant speech. Methods that only use a reverberant speech signal for estimating the reverberation time are commonly denoted *blind* reverberation time estimators. In [9], a time-frequency model that takes the interaction between the decay rates of the room and speech into account is used, and in [10] a multilayer feed forward neural network is used. An extension of the work in [9] has recently been presented in [11], showing increased noise robustness. Another noise-robust approach, operating in the short time Fourier transform (STFT) domain for reduced computational complexity, has also recently been proposed in [12]. In [13], an adaptive measure, termed speech to reverberation modulation energy ratio, has been developed using modulation spectral insights; it is shown that the inverse of this ratio is highly correlated with the reverberation time. Further, a combination of the methods in [9] and [13] has been proposed in [14], using spectral decomposition of the speech signal to allow T_{60} estimation in different frequency sub-bands. All six methods [9], [10], [11], [12], [13], [14] are blind, but rely on prior training for parameter calibration.

Another approach to blind reverberation time estimation, not relying on prior training, is presented in [7], where sections of the microphone signal where the energy of the reverberated signal decays

exponentially is detected, and then the slope of the logarithm of the smoothed energy envelope (corresponding to the reverberation time) is estimated through linear regression. An alternative approach based on *maximum-likelihood estimation* (ML estimation) was presented in [15], where the diffusive reverberation tail is modeled as an exponentially damped Gaussian white noise process. The method continuously estimates the reverberation time and an order-statistics filter was used to extract the most likely reverberation time from the accumulated estimates, hence avoiding the need of explicitly finding sections of the signal with decaying energy. Efforts for reducing the computational complexity of the method in [15] have also been made in [16] and [17] as well as a related approach for improved noise robustness in [18], at the cost of increased computational complexity. A recent approach presented in [19] introduces a pre-processing step involving a linear prediction filter for signal whitening, which was shown to improve estimation robustness.

The methods in [9], [13] and [17] have recently been compared in an experimental study in [20] and it was concluded that all of the methods are able to estimate the reverberation time to within ± 0.2 s for $T_{60} < 0.8$ s and signal-to-noise ratio (SNR) ≥ 30 dB; and that the method in [17] requires the least computational complexity. It was also shown that additive noise causes positive estimation error bias for all methods, which shows the importance of noise robust solutions.

The general blind reverberation time estimator comprises multiple algorithms/processing blocks. For example, the approach in [17] involves signal downsampling, an elaborate method for finding sections of the signal with decaying energy, estimation of the decay rate of the respective decaying energy signal sections, and finally, a two-step histogram based order-statistics filtering process. The method in [15] uses neither downsampling nor any algorithm for finding energy decay segments, but still uses the same decay rate estimator together with a similar histogram based order-statistics filter. The related method described in [19] uses a whitening pre-processor, combined with the decay rate estimator, but without any filtering for post-processing of the estimates.

It should be emphasized that the methods in [15], [16], [17] and [19] rely on *the same* maximum-likelihood (ML) based decay-rate estimation approach. As a comparison, the methods in [7], [9], [11], [12] and [14] are all based on a linear-regression (LR) type decay rate estimation. Despite this clear division of blind T_{60} estimation methods; using either ML- or LR-type decay rate estimators, there has been no comparison between ML and LR in the literature, to the best of our knowledge. The study in [20] was purely experimental (using real recorded speech signals) and compared the “complete” T_{60} estimators, including possible segmentation and smoothing (i.e., not just the decay rate estimation part). Furthermore, both ML- and LR-based decay rate estimators are based upon a noiseless assumption, or

that the noise is “negligible”, and no theoretical analysis of the influence of noise on these methods has been conducted.

This paper presents an in-depth comparison between the ML- and LR-type decay rate estimators and their performance under the influence of additive noise is investigated. Comparison is also made with the noise robust estimation approach in [18], here denoted MLN. Further, a hybrid approach to significantly reduce the computational complexity of MLN is presented. The key contributions of the paper are as follows:

- In a noiseless scenario with an i.i.d. Gaussian excitation signal, it is shown that the LR-based decay rate estimator is unbiased if no smoothing is used, and a closed form expression for the estimation variance is derived. It is shown that the estimation variance is about 4 dB above the Cramér-Rao lower bound (i.e., the lowest possible estimation variance achievable by any unbiased parametric estimator).
- The performance of the LR-, ML- and MLN approaches are analyzed in the case of i.i.d. Gaussian background noise. An analytic expression for the estimation bias is obtained and it is argued, and verified through Monte-Carlo simulations that this estimation bias is similar for both LR and ML-based estimators, and that the noise induced increase in mean-square error is mostly caused by this bias.
- A hybrid approach, combining the noise robust performance of MLN with the lower computational complexity of ML is proposed, and it is shown that the estimation performance is indeed similar to MLN, while having computational complexity similar to ML.

Experiments with real speech signals are intentionally excluded, due to the fact that this would require either automatic or manual detection and segmentation of reverberant speech pauses, which in turn would affect the estimation performance. Methods for automatic detection of speech pauses have been presented in, e.g., [7], [17], and could of course be used together with LR, ML, MLN or any other decay rate estimator. It should be recognized that the objective here is to, given an observed sequence according to the diffusive reverberation tail model (see section II), study the estimation performance of decay rate estimators; an important component in many state-of-the art blind T_{60} estimators.

The paper is organized as follows. In section II, the diffusive reverberation tail model is presented. In section III and IV, respectively, the LR based approach and the ML based approach are analyzed in terms of performance with and without interfering additive noise. The noise robust approach MLN is explained in section V. Approximate analytical expressions for the bias and MSE are presented for

all three methods. Then, in subsection V-B, a hybrid ML-MLN approach is presented. Monte-Carlo simulations to verify the results are presented in section VI and section VII discusses the results with respect to blind T_{60} estimation. Finally, conclusions are in section VIII.

II. DIFFUSIVE REVERBERATION TAIL MODEL

In this section, the diffusive reverberation tail assumption is used to form two models, which in the remainder of the text are used for both derivation as well as evaluation of decay rate estimators. The first model is based on a noiseless assumption, while the second also considers additive noise.

A. Noiseless assumption

Assuming a source-microphone distance greater than the critical distance and a diffuse sound field [21], means that the beginning of a reverberant speech pause can be modeled the same as in [5], [9], [15], [16], [17]

$$s(n) = v(n)e^{-\rho n}, \quad (1)$$

where $v(n)$ is a random signal, typically described as a stationary i.i.d. white Gaussian sequence, and $\rho > 0$ determines the decay rate. Approaches where $v(n)$ is assumed to be Laplacian have also been presented in the literature [22], [23], although the Gaussian assumption is most common. Even though speech is typically non-white, it has been shown that using this simple model still works rather well (e.g., [15]). Moreover, it is also possible to use pre-whitening [19], [24] for even more accurate model matching.

The instantaneous power

$$E\{s^2(n)\} = E\{v^2(n)\}e^{-2\rho n}, \quad (2)$$

where $E\{\cdot\}$ denotes expectation, gives the relation between ρ and the reverberation time as

$$T_{60} = \frac{3}{\rho F_s \log_{10}(e)} \approx \frac{6.91}{\rho F_s}, \quad (3)$$

where F_s is the sampling frequency.

In [15], the Cramér-Rao lower bound (CRB), which expresses the lower bound for the variance of a deterministic parameter estimator [25], was calculated based upon the previously described model and the derivation is briefly recited below (in slightly different notation). Note that the CRB in [15] was calculated for the estimation of a , where $a = e^{-\rho}$, while in the following the CRB for the estimation of ρ is considered. Assuming, as mentioned previously, that $v(n)$ is i.i.d. zero mean Gaussian with variance σ_v^2

gives the joint probability density function (likelihood function) for N observations ($n = \{0, \dots, N-1\}$) of $s(n)$ as

$$\mathcal{L}(\rho, \sigma_v) = \left(\frac{1}{2\pi\sigma_v^2 e^{-(N-1)\rho}} \right)^{N/2} \times \exp \left(-\frac{1}{2\sigma_v^2} \sum_{n=0}^{N-1} e^{2n\rho} s^2(n) \right). \quad (4)$$

By taking the logarithm of (4), followed by the 2nd partial derivatives with respect to ρ and σ_v , allows the calculation of the Fisher information matrix, which in this case is defined as

$$\mathbf{J} = -E \left\{ \begin{pmatrix} \frac{\partial^2 \ln \mathcal{L}(\rho, \sigma_v)}{\partial \rho^2} & \frac{\partial^2 \ln \mathcal{L}(\rho, \sigma_v)}{\partial \rho \partial \sigma_v} \\ \frac{\partial^2 \ln \mathcal{L}(\rho, \sigma_v)}{\partial \sigma_v \partial \rho} & \frac{\partial^2 \ln \mathcal{L}(\rho, \sigma_v)}{\partial \sigma_v^2} \end{pmatrix} \right\}, \quad (5)$$

yielding

$$\mathbf{J} = \begin{pmatrix} \frac{1}{3}N(N-1)(2N-1) & \frac{1}{\sigma_v}N(N-1) \\ \frac{1}{\sigma_v}N(N-1) & \frac{2N}{\sigma_v^2} \end{pmatrix}. \quad (6)$$

The bound on the covariance matrix of the estimates of ρ and σ_v are then obtained through the inverse of (6), i.e.,

$$\mathbf{J}^{-1} = \begin{pmatrix} \frac{6}{N(N^2-1)} & -\frac{3\sigma_v}{N(N+1)} \\ -\frac{3\sigma_v}{N(N+1)} & \frac{\sigma_v^2(2N-1)}{N(N+1)} \end{pmatrix}, \quad (7)$$

which gives

$$\text{CRB}(\rho) = \frac{6}{N(N^2-1)}, \quad (8)$$

and

$$\text{CRB}(\sigma_v) = \frac{\sigma_v^2(2N-1)}{N(N+1)}. \quad (9)$$

As the focus in this paper is the estimation of the decay rate ρ , the noiseless CRB, i.e., (8), is here used for benchmarking the estimation performance of all considered estimators.

B. Additive noise

In a natural environment, it is reasonable to assume that there is additive noise, and the model in (1) can then be extended as

$$s(n) = v(n)e^{-\rho n} + d(n), \quad (10)$$

where $d(n)$ is the interfering noise signal. The models in (1) and (10) are much related to that assumed in, e.g., backward integration of an RIR [26]. The difference is of course that in RIR backward integration,

the RIR *itself* is modeled by an exponential decay; whereas in the case considered here, the reverberated *signal* is modeled by an exponential decay. It should, however, be noted that the fundamental concept of backward integration has also been used for blind approaches, see e.g., [14], [27]. Backward integration in this context is based on the basic relation that

$$\int_t^\infty e^{-\rho\tau} d\tau = \frac{1}{\rho} e^{-\rho t}, \quad (11)$$

i.e., the decay rate is preserved while the integration provides smoothing to reduce variance. (It was also shown in [28] that the RIR backward integration is mathematically equivalent to the ensemble average of decay curves of multiple interrupted white noise trials.) In practice, the RIR backward integration is typically calculated (see e.g., [14] and [29]), as

$$c(n) = \frac{\sum_{k=n}^{L-1} h^2(k)}{\sum_{k=0}^{L-1} h^2(k)}, \quad (12)$$

where $h(k)$ is the measured impulse response of length L with assumed exponential decay and $n = \{0, \dots, L-1\}$. Then, linear (or non-linear) regression is used to find the slope of $\ln c(n)$. Unfortunately, there are a number of issues with this approach. First of all, the tail of $c(n)$ will be bent down as n approaches L (reaching zero ($-\infty$ dB) for $n = L-1$), distorting the desired exponential shape. Only considering $c(n)$ for n up to a certain index is often used in practice, but finding the optimal range is non-trivial [26]. Increasing L can be beneficial, but only up to a certain point; the noise floor will at some point begin shifting $c(n)$ away from the exponential decay, resulting in a positive bias. These effects (and more) have been investigated in [26] and some rules-of-thumb were presented; specifically to truncate the measured impulse response at the knee where the main decay slope intersects the noise floor, then measure the slope of the backward integrated truncated impulse response down to a level about 5 dB above the noise floor. However, methods for, e.g., finding the knee were not discussed.

III. LINEAR REGRESSION BASED DECAY RATE ESTIMATION

A straight-forward approach for estimating ρ , assuming the model in (1), is simply to take the natural logarithm of each squared sample observed and then estimate ρ through linear regression [7]. This approach is used in many state-of-the art blind reverberation time estimators (e.g., [9], [11], [12], [14]). The method is described in detail below.

Squaring and taking the natural logarithm of (1) yields

$$z(n) = w(n) - 2\rho n, \quad (13)$$

where $z(n) = \ln s^2(n)$ and $w(n) = \ln v^2(n)$. A linear regression model can be written as

$$\hat{z}(n) = \hat{w} - 2\hat{\rho}_{\text{LR}}n, \quad (14)$$

where $\hat{w}$ and $\hat{\rho}_{\text{LR}}$ are the two model parameters. For a set S of observations, the linear least-square error (LLSE) is written as

$$\sum_{n \in S} (z(n) - \hat{w} + 2\hat{\rho}_{\text{LR}}n)^2. \quad (15)$$

In order to find the $\hat{\rho}_{\text{LR}}$ which minimizes the LLSE, the derivative of (15) with respect to $\hat{\rho}_{\text{LR}}$ is set to 0, i.e.,

$$4 \sum_{n \in S} n (z(n) - \hat{w} + 2\hat{\rho}_{\text{LR}}n) = 0, \quad (16)$$

which can be re-written as

$$\sum_{n \in S} n (z(n) + 2\hat{\rho}_{\text{LR}}n) - \hat{w} \sum_{n \in S} n = 0. \quad (17)$$

It can now be seen that if the elements in S are symmetrical around 0, then $\sum_{n \in S} n = 0$, and the LLSE estimate of ρ is given by

$$\hat{\rho}_{\text{LR}} = -\frac{\sum_{n \in S} n z(n)}{2 \sum_{n \in S} n^2}. \quad (18)$$

Worth noting is that (18) is a standard linear regression approach, with a regressor vector constituting the elements in S . Even though the model in (1) was assumed in the beginning of this section, the approach could thus be used for estimating ρ as long as the observed samples of $z(n)$ are approximately in the mean linearly decreasing with a factor 2ρ , regardless of the exact underlying model and any i.i.d. Gaussianity assumption(s). This should be kept in mind when reading the part in the following subsection where smoothing of $s(n)$ before taking the logarithm is considered.

A. Estimation bias

It can easily be seen that the estimator is unbiased as

$$\begin{aligned} E\{\hat{\rho}_{\text{LR}} - \rho\} &= -\frac{\sum_{n \in S} n E\{z(n)\}}{2 \sum_{n \in S} n^2} \\ &\quad - \frac{\sum_{n \in S} n (E\{w(n)\} - E\{z(n)\})}{2 \sum_{n \in S} n^2} \\ &= -E\{w(n)\} \frac{\sum_{n \in S} n}{2 \sum_{n \in S} n^2} \\ &= 0. \end{aligned} \quad (19)$$

However, the original approach as presented in [7] used recursive smoothing of $s^2(n)$, i.e.,

$$\bar{s}(n) = \beta \bar{s}(n-1) + (1-\beta)s^2(n), \quad (20)$$

where $0 \leq \beta < 1$ is a forgetting factor, before taking the logarithm and then estimating through linear regression. Inserting (1) and taking the expected value (ensemble average) of $\bar{s}(n)$ gives

$$\begin{aligned} E\{\bar{s}(n)\} &= E\{v^2(n)\}e^{-2\rho n}(1-\beta) \\ &\times \left(\sum_{i=n_0}^n \beta^{i-n_0} e^{2\rho(i-n_0)} \right), \end{aligned} \quad (21)$$

where n_0 is the first element in S . The bias in this case will then be

$$E\{\hat{\rho}_{\text{SLR}} - \rho\} = - \frac{\sum_{n \in S} n \ln \left(\sum_{i=n_0}^n \beta^{i-n_0} e^{2\rho(i-n_0)} \right)}{2 \sum_{n \in S} n^2}. \quad (22)$$

It can be seen from (22) that increasing β , i.e., the smoothing, will increase the bias, while increasing N will decrease the bias. With this in mind, no smoothing, i.e., $\beta = 0$, is assumed in the following text.

In the case of backward integration (see section II), the smoothed signal is modeled (with the sum approximated by an integral for the sake of mathematical tractability) as

$$c_{\text{BLR}}(n) \simeq \frac{1}{\eta} \int_n^{n_B} v^2(k) e^{-2\rho\tau} d\tau, \quad (23)$$

where $\simeq$ denotes an equality where only the dominant terms have been retained, $\eta = \sum_{k \in S} v^2(k) e^{-2\rho k}$ and n_B is the backward integration “starting element”. Inserting (1) into (23) and taking the expected value yields

$$\begin{aligned} E\{c_{\text{BLR}}(n)\} &\simeq E\{v^2(n)\}e^{-2\rho n} \left(1 - e^{-2\rho(n_B-n)} \right) \\ &\times E\left\{ \frac{1}{2\rho\eta} \right\}. \end{aligned} \quad (24)$$

Taking the logarithm of (24) and estimating ρ using the previously described linear regressor will then give the bias as

$$E\{\hat{\rho}_{\text{BLR}} - \rho\} \simeq - \frac{\sum_{n \in S} n \ln \left(1 - e^{-2\rho(n_B-n)} \right)}{2 \sum_{n \in S} n^2}, \quad (25)$$

and it can clearly be seen that the larger n_B is, compared to n , the less bias. It might be tempting to approximate the logarithm in (25) by a low order Taylor expansion; using, e.g., $\ln(1 - e^{-x}) \approx \ln x$ or $\ln(1 + y) \approx y$ (where in this case $y = -e^{-x}$) for further simplification. However, this will only yield accurate results for comparatively large ρ and n_B , since otherwise $e^{-2\rho(n_B-n)}$ will be very close to 1 for large n (close to n_B), which in turn implies significant Taylor expansion approximation error. However, for comparatively large ρ and n_B , the estimation bias will be close to 0, and thus, an analytical expression valid only in this case is not particularly interesting.

Nevertheless, the similarity between (25) and the expression for the bias in the case of interfering additive noise given by (36) should be noted; the backward integration will impose a positive bias and any interfering additive noise will impose a negative bias. By choosing n_B appropriately, an unbiased estimator will be obtained. This is also what is shown in [26]. However, tuning n_B is non-trivial [26]. In the following text, no backward integration is assumed.

B. Estimation variance

The variance of the estimator in (18) is given by

$$E\{(\hat{\rho}_{\text{LR}} - \rho)^2\} = \frac{\sum_{n \in S} \sum_{k \in S} nk E\{w(n)w(k)\}}{(2 \sum_{n \in S} n^2)^2}. \quad (26)$$

Further, assuming the observed set $S = \{-(N-1)/2, \dots, 0, \dots, (N-1)/2\}$ of length N and uncorrelated samples, i.e., $E\{w(n)w(k)\} = E\{w(n)\}E\{w(k)\} \quad \forall k \neq n$, (26) can be re-written as

$$\begin{aligned} E\{(\hat{\rho}_{\text{LR}} - \rho)^2\} &= \frac{36 (2E\{w^2(n)\} - E\{w(n)\}^2)}{N^2(N^2 - 1)^2} \\ &\times \sum_{n=0}^{(N-1)/2} n^2 \\ &= \frac{3 (E\{w^2(n)\} - E\{w(n)\}^2)}{N(N^2 - 1)} \end{aligned} \quad (27)$$

By defining $v(n) = \sigma_v n(n)$, it can be seen that $\ln v^2(n) = \ln \sigma_v^2 + \ln n^2(n)$, and thus, $E\{(\ln v^2(n))^2\} - E\{(\ln v^2(n))\}^2 = E\{(\ln n^2(n))^2\} - E\{(\ln n^2(n))\}^2$, i.e., the difference $E\{w^2(n)\} - E\{w(n)\}^2$ does not depend on the variance of $v(n)$. Assuming that $v(n)$ is i.i.d. Gaussian gives that $v^2(n)$ has a chi-square distribution with one degree of freedom. The mean and variance of $w(n)$ are then [30]

$$\mu_w = E\{w(n)\} = \psi\left(\frac{1}{2}\right) - \ln\left(\frac{1}{2}\right), \quad (28)$$

$$E\{(w(n) - \mu_w)^2\} = \psi'\left(\frac{1}{2}\right), \quad (29)$$

where $\psi(z)$ is the digamma function and $\psi'(z) = \frac{\partial}{\partial z} \psi(z)$. Thus, the variance of the estimator is given by

$$E\{(\hat{\rho}_{\text{LR}} - \rho)^2\} = \frac{3\pi^2}{2N(N^2 - 1)}. \quad (30)$$

From (30), it can be seen that the variance of the linear regression based estimator is $(\pi/2)^2$ times (that is, about 4 dB) that of the CRB (8), regardless of ρ and N .

C. Robustness to additive noise

The same estimator as described earlier is considered, but now in a noisy scenario, given by (10), which can be re-written as

$$s(n) = (v(n) + d(n)e^{\rho n}) e^{-\rho n}. \quad (31)$$

Squaring and taking the logarithm, as before, yields

$$\ln s^2(n) = u(n) - 2\rho n, \quad (32)$$

where $u(n) = \ln(v(n) + d(n)e^{\rho n})^2$. Assuming that $v(n)$ and $d(n)$ both are uncorrelated zero mean i.i.d. Gaussian with variances σ_v^2 and σ_d^2 , respectively, gives that the sum is also a zero mean i.i.d. Gaussian. The expression $v(n) + d(n)e^{\rho n}$ is thus a zero mean Gaussian with variance $\sigma_v^2 + e^{2\rho n}\sigma_d^2$ and $(v(n) + d(n)e^{\rho n})^2$ has a chi-square distribution. The mean of $u(n)$ is, as in the previous section,

$$E\{u(n)\} = \psi\left(\frac{1}{2}\right) - \ln\left(\frac{1}{2}\right) + \ln(\sigma_v^2 + e^{2\rho n}\sigma_d^2), \quad (33)$$

which is re-written as

$$E\{u(n)\} = -\gamma - \ln(\sigma_v^2/2) + \ln\left(1 + e^{2\rho n + \ln(\sigma_d^2/\sigma_v^2)}\right), \quad (34)$$

where γ is the Euler–Mascheroni constant. In order to obtain an expression for the bias, i.e.,

$$b(\rho) = -\frac{\sum_{n \in S} n E\{u(n)\}}{2 \sum_{n \in S} n^2}, \quad (35)$$

the sum in (35) is approximated by an integral

$$b(\rho) \simeq -\frac{1}{2 \sum_{n \in S} n^2} \int_{-(N-1)/2}^{(N-1)/2} x \ln\left(1 + e^{2\rho x + \nu}\right) dx, \quad (36)$$

where $\nu = \ln(\sigma_d^2/\sigma_v^2) + \rho(N-1)$ and the constant terms of $E\{u(n)\}$ disappear owing to symmetry. (The $\rho(N-1)$ term is for correction of the SNR in relation to the index n ; since σ_d^2/σ_v^2 is the inverse SNR at $n=0$ (prior to adding the correction term), while the observed sequence is taken between $-(N-1)/2$ and $(N-1)/2$, the correction term is added so that σ_d^2/σ_v^2 will represent the inverse SNR at the first sample of the sequence, i.e., $n = -(N-1)/2$.)

Using partial integration and substitution (see appendix), equation (36) can be written as

$$b(\rho) \simeq \frac{3}{2} \left(\frac{\text{Li}_2(\beta_2) + \text{Li}_2(\beta_1)}{N(N+1)\rho} + \frac{\text{Li}_3(\beta_2) - \text{Li}_3(\beta_1)}{N(N^2-1)\rho^2} \right). \quad (37)$$

where $\beta_1 = -(\sigma_d^2/\sigma_v^2) \exp(2\rho(N-1))$, $\beta_2 = -\sigma_d^2/\sigma_v^2$ and $\text{Li}_s(\cdot)$ is the Polylogarithm function of order s . It can be seen that since $\text{Li}_s(0) = 0$, the expression in (37) will be 0 for $\sigma_d^2 = 0$, i.e., the estimator is unbiased in the noiseless case (as shown previously).

In the noisy case, when the SNR is “large enough”, the impact of the noise is negligible. It can then be argued that as the SNR decreases, the impact of the noise will primarily manifest itself in the estimation bias and that the variance of the estimator is, in comparison, not affected by the noise. Thus, the mean square error of the estimator in the noisy case can be approximated as

$$E\{(\hat{\rho}_{\text{LR}} - \rho)^2\} + b^2(\rho), \quad (38)$$

where the variance was given by (30) and the bias in (37). It can easily be realized that for $N \rightarrow \infty$, the estimate $\hat{\rho}_{\text{LR}}$ will go to 0 as the noise $d(n)$ (which has constant variance), will be totally dominating. This can also be seen analytically using the limit [31]

$$\lim_{\text{Re}(\omega) \rightarrow \infty} \text{Li}_s(-e^\omega) = -\frac{\omega^s}{\Gamma(s+1)}, \quad (39)$$

where $\Gamma(z)$ is the gamma function. Combining (39) and (37), gives

$$\lim_{N \rightarrow \infty} b(\rho) \simeq \frac{3}{2} \left(-\frac{(2\rho(N-1) + \beta_3)^2}{2N(N+1)\rho} + \frac{(2\rho(N-1) + \beta_3)^3}{6N(N^2-1)\rho^2} \right), \quad (40)$$

where $\beta_3 = \ln(\sigma_d^2/\sigma_v^2)$, which after some algebraic manipulation reduces to

$$\lim_{N \rightarrow \infty} b(\rho) = -\rho, \quad (41)$$

meaning that, as stated above, that the estimator will be totally dominated by the noise.

IV. MAXIMUM-LIKELIHOOD BASED DECAY RATE ESTIMATION

In this section, the method from [15] (also used in, e.g., [16], [17], [19]) is described (in slightly different notation). With the same i.i.d. Gaussian assumption as in section II, the likelihood function for N observations of $v(n)$ is given by (4), and the log-likelihood function is

$$\ln \mathcal{L}(\rho, \sigma_v) = \frac{N}{2} \left((N-1)\rho - \ln(2\pi\sigma_v^2) \right) - \frac{1}{2\sigma_v^2} \sum_{n=0}^{N-1} e^{2n\rho} s^2(n). \quad (42)$$

Taking the partial derivatives of (42), with respect to ρ and σ_v yields

$$\frac{\partial \ln \mathcal{L}(\rho, \sigma_v)}{\partial \rho} = \frac{N(N-1)}{2} - \frac{1}{\sigma_v^2} \sum_{n=0}^{N-1} n e^{2n\rho} s^2(n), \quad (43a)$$

$$\frac{\partial \ln \mathcal{L}(\rho, \sigma_v)}{\partial \sigma_v} = -\frac{N}{\sigma_v} + \frac{1}{\sigma_v^3} \sum_{n=0}^{N-1} e^{2n\rho} s^2(n). \quad (43b)$$

Setting (43a) and (43b) to zero and combining them gives

$$\frac{\sum_{n=0}^{N-1} n e^{2\rho n} s^2(n)}{\sum_{n=0}^{N-1} e^{2\rho n} s^2(n)} = \frac{N-1}{2}, \quad (44)$$

which can be solved numerically for ρ in order to obtain the maximum. The asymptotic properties of ML estimators [25] means that the considered ML approach is asymptotically unbiased and attains the CRB (8).

In [15], it was proposed to first use a bisection method and then Newton-Raphson, requiring around 10 – 12 iterations in total to obtain a root for (44) with reasonable precision. Two faster methods were presented in [16]; one based on a recursive technique for calculating consecutive solutions for ρ and the other an iterative solution (related to Newton-Raphson) relying on the tuning of a step-size parameter. With the appropriate step-size parameter tuning, the method required approximately 5 iterations. In [17], a quantization approach was used where the log-likelihood function in (42) was evaluated for a fixed number of values of ρ , and the ρ yielding the maximum log-likelihood was selected as the root. Another approach was taken in [23], based on the assumption of $|v(n)|$ being Exponentially distributed (as opposed to $v(n)$ being Gaussian). The assumption yields a similar log-likelihood function as (42), but with a somewhat reduced steepness. This reduced steepness facilitates basic fitting of a 2nd degree polynomial curve to the log-likelihood function, reducing the overall complexity required for finding the maximum log-likelihood, but without sacrificing significant estimation performance.

A. Robustness to additive noise

The same additive noise as in section III-C, i.e., the noisy scenario given by (10), is assumed. Since the properties of both the ML and the LR estimators were shown to be similar in the noiseless case (unbiased and the variance of the LR estimator is that of the ML estimator (i.e., the CRB (8)) times a constant of approximately 2.5), it can also be argued that the bias in the noisy case for both estimators is fairly similar. Also, using the assumption that noise will primarily affect the estimation bias rather than the variance, i.e., the same assumption as for the linear regression case, suggests that the mean square error of the ML estimator can be approximated as

$$E\{(\hat{\rho}_{\text{ML}} - \rho)^2\} + b^2(\rho), \quad (45)$$

where the variance is given by the CRB in (8) and the bias in (37).

A more accurate expression for the mean square error can be obtained using the general expression for the multivariate CRB [25], given by

$$\mathbf{C} \geq \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{J}^{-1} \frac{\partial g(\boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}}, \quad (46)$$

where $\mathbf{C}$ is the estimation error covariance matrix, $\boldsymbol{\theta} = [\rho, \sigma_v]^T$, $\mathbf{J}$ is the Fisher information matrix as in (6) and $g(\boldsymbol{\theta}) = b(\boldsymbol{\theta}) + \boldsymbol{\theta}$, where $b(\boldsymbol{\theta})$ is the estimation bias. Naturally, this would also require an expression for the σ_v estimation bias (in addition to the expression for the ρ estimation bias given by (37)). However, as will be shown in section VI, the MSE approximation given by (45) is fairly accurate.

V. MAXIMUM-LIKELIHOOD BASED DECAY RATE ESTIMATION WITH NOISE TAKEN INTO ACCOUNT

The LR and ML estimators described in sections III and IV, respectively, do not take any additive noise into account and thus suffer from estimation bias if the SNR is not “large enough”. An ML-based estimator that does take the additive noise into account was presented in [18], and is described below.

The observed signal is modeled as in (10). Assuming that $v(n)$ and $d(n)$ are two uncorrelated i.i.d. Gaussian signals with zero mean and with variances σ_v^2 and σ_d^2 , respectively, give the likelihood function for N observations ($n = \{0, \dots, N-1\}$) of $s(n)$ as

$$\mathcal{L}_N(\rho, \sigma_v) = \left(\frac{1}{(2\pi)^N \prod_{n=0}^{N-1} \sigma_v^2 e^{-2\rho n} + \sigma_d^2} \right)^{1/2} \times \exp \left(-\frac{1}{2} \sum_{n=0}^{N-1} \frac{s^2(n)}{\sigma_v^2 e^{-2\rho n} + \sigma_d^2} \right), \quad (47)$$

and the log-likelihood function as

$$\ln \mathcal{L}_N(\rho, \sigma_v) = -\frac{1}{2} \left(\ln(2\pi) + \sum_{n=0}^{N-1} \ln(\sigma_v^2 e^{-2\rho n} + \sigma_d^2) + \sum_{n=0}^{N-1} \frac{s^2(n)}{\sigma_v^2 e^{-2\rho n} + \sigma_d^2} \right). \quad (48)$$

The variance σ_d^2 is assumed to be known, since it can normally be estimated directly in the absence of speech [32]. In order to find the maximum of (48), the partial derivatives with respect to ρ and σ_v^2 are set to zero, yielding

$$\sum_{n=0}^{N-1} \frac{n\sigma_v^2 e^{-2\rho n}}{\sigma_v^2 e^{-2\rho n} + \sigma_d^2} \left(1 - \frac{s^2(n)}{\sigma_v^2 e^{-2\rho n} + \sigma_d^2} \right) = 0, \quad (49a)$$

$$\sum_{n=0}^{N-1} \frac{\sigma_v^2 e^{-2\rho n}}{\sigma_v^2 e^{-2\rho n} + \sigma_d^2} \left(1 - \frac{s^2(n)}{\sigma_v^2 e^{-2\rho n} + \sigma_d^2} \right) = 0. \quad (49b)$$

In [18], it was proposed to solve the corresponding equations (49a) and (49b) through *expectation-maximization*. This means that, e.g., a value of σ_v^2 is picked arbitrarily and kept fixed, while an estimate of ρ is obtained using (49a) and then the estimated ρ is kept fixed while σ_v^2 is estimated using (49b), and this iterating process is continued until convergence has been reached. Alternatively, (48) is maximized with

a numerical method (bisection and/or Newton-Rhapson) using the same iterative approach just described (ρ and σ_v^2 are each alternately estimated while the other is kept fixed, until convergence is reached).

Another approach, for reduced computational complexity, was also presented in [18], where σ_v^2 could be estimated as

$$\hat{\sigma}_v^2 = \frac{1}{N_L} \sum_{n=0}^{N_L-1} s^2(n), \quad (50)$$

if the interfering noise is not too strong and N_L covers a period of about 20 ms or less so that the sound decay has no significant influence. Then ρ can be calculated directly using (49a) in a *non-iterating* approach.

A. Estimation variance

In the noiseless case, the $\hat{\rho}_{ML}$ estimation variance will go to 0 as $N \rightarrow \infty$, as shown by (8). In the case of interfering noise, however, this noise will impose a lower bound on the estimation variance - once the reverberation tail is, in a sense, overpowered by the noise, increasing N will not improve estimation accuracy. This can be realized by, e.g., considering the problem of estimating the decay rate of a (deterministic) damped (i.e., decaying) sinusoid in constant additive noise; it has been shown that contrary to the undamped case, the CRB of a damped (deterministic) process will not converge to zero when N increases [33]. It can thus be argued that this is also true for a damped random process. A basic model of the estimation variance is: as long as all samples have positive SNR, the variance attains the CRB for the noiseless case, i.e., (8), and samples where the SNR is negative do not contribute to reducing the variance. The sample index corresponding to 0 dB is given by (using (3))

$$N_{0dB} = \frac{\ln(\sigma_v^2/\sigma_d^2)}{2\rho}, \quad (51)$$

and then simply adding the corresponding variance to that of the noiseless case yields

$$E\{(\hat{\rho}_{MLN} - \rho)^2\} \simeq \frac{6}{N(N^2 - 1)} + \frac{6}{N_{0dB}(N_{0dB}^2 - 1)}. \quad (52)$$

The accuracy of (52) is shown through simulations in section VI.

B. Proposed hybrid approach

A better estimate of σ_v^2 than that obtained by (50) can be achieved using (43a) set to zero, i.e.,

$$\hat{\sigma}_v^2 = \frac{1}{N_L} \sum_{n=0}^{N_L-1} e^{2\rho n} s^2(n). \quad (53)$$

This estimate will be accurate as long as the SNR is not too poor and N_L is not too large in relation to the decay rate ρ . For example, by assuming a SNR of at least 20 dB, and that $\rho \leq 0.008$, which

corresponds to a minimum reverberation time of approximately 0.1 seconds (see (3)) for a sampling frequency of 8 kHz, it can be seen that in this “worst case” scenario, the reverberation tail will have dropped down to the noise floor after around 287 samples. Choosing N_L below this value will make sure that the estimate in (53) is never significantly affected by the interfering noise. By examining (8) and (9), i.e., the CRB of the ρ and σ_v estimates in the noiseless case, it can clearly be seen that the lower bound for ρ decreases with $1/N^3$ while the lower bound for σ_v decreases with $1/N$. Thus, using fewer samples for estimating σ_v would not have as negative of an effect on the estimation variance as using fewer samples for estimating ρ .

Furthermore, combining (53) with (49a) will also yield a *non-iterating* approach, which saves a significant amount of computational resources. Of course, the computational complexity depends on the implementation. If an iterative method as in [15] or [16] is used, then the number of multiplications and divisions required for the proposed approach is as follows. By writing (48) as

$$\ln \mathcal{L}_N(\rho, \sigma_v) = -\frac{1}{2} \left(\ln(2\pi) + \sum_{n=0}^{N-1} \left(\ln \sigma_v^2 + \ln \left(e^{-2n\rho} + \sigma_d^2 / \sigma_v^2 \right) \right) + \frac{1}{\sigma_v^2} \sum_{n=0}^{N-1} \frac{s^2(n)}{e^{-2n\rho} + \sigma_d^2 / \sigma_v^2} \right), \quad (54)$$

it can be seen that N multiplications are required for calculating $e^{-2n\rho}$, N multiplications are needed for $s^2(n)$, N divisions are required for evaluating the terms of the second sum and the rest of the expression can be calculated using 3 additional multiplications/divisions. In addition, (53) needs to be calculated, requiring a total of N_L multiplications (for evaluating $e^{2n\rho}$) and 1 additional multiplication/division. An iterative approach using R iterations, i.e., evaluating both (54) and (53) R times, would then require a total of $R(3(N+1) + N_L + 1)$ multiplications/divisions. Table I shows the computational complexity (in number of multiplications and divisions per estimate) of LR, ML (the fast approach in [16]), and MLN, as well as hybrid MLN. In the table, M is the number of expectation-maximum iterations for the MLN method. Setting for example $R = 5$ (as in [16]), $M = 4$ (as in [18]), $N_L = 250$ and $N = 1000$, gives a total of 16270 multiplications/divisions for the proposed method (similar to the 16010 of ML), while MLN requires 120120, i.e., more than 7 times as many.

VI. SIMULATIONS AND DISCUSSION

The expressions for the estimation bias and mean square error of all four estimators, summarized in Table II, were verified through Monte-Carlo simulations and the ensemble average over 1000 simulations was considered for each parameter setting. It should be noted that it is the decay rate estimation

TABLE I
COMPUTATIONAL COMPLEXITY (IN NUMBER OF MULTIPLICATIONS AND DIVISIONS PER ESTIMATE) OF THE COMPARED ESTIMATORS

Estimator	Number of mult. & div.
LR [7]	$2N + 1$
ML (fast approach in [16])	$R(3N + 2) + N$
MLN (direct approach [18])	$6MR(N + 1)$
Proposed method (Hybrid ML)	$R(3(N + 1) + N_L + 1)$

TABLE II
THE COMPARED ESTIMATORS WITH CORRESPONDING EXPRESSIONS FOR BIAS AND VARIANCE

Estimator	Bias	Variance
LR [7]	(37)	(30)
ML [15]	(37)	(8)
MLN [18]	-	(52)
Hybrid ML (Proposed method)	-	(52)

components (see section I) first proposed in [7], [15] and [18] that are compared, and *not* the complete blind reverberation time estimation approaches in the respective reference.

The initial SNR at sample 0 was set to 20 dB and decreased as the sample index increased due to the exponential decay, illustrated in Fig. 1. Four different decay rates were considered, $\rho = \{0.008, 0.004, 0.002, 0.001\}$, corresponding to T_{60} reverberation times of approximately 0.1, 0.2, 0.4 and 0.8 seconds, for a sample rate F_s of 8 kHz. The number of expectation-maximization iterations for MLN was set to 16 and the parameter $N_L = \min\{250, N\}$ (slightly below 287 to ensure a positive SNR with some margin) was used for the proposed hybrid MLN.

In Fig. 2, the theoretical and simulated bias of the LR (upper plot) and ML (lower plot) estimators are shown for different values of ρ and N . The figure clearly shows that the estimation bias of both methods is essentially the same and that the theoretical expression for the bias is indeed accurate. It should be recognized that there is some bias even for N where the SNR is fairly high (e.g., $N < 200$). This indicates the importance of considering the interfering noise in such situations.

The theoretical and simulated mean square error of the different estimators are shown in Fig. 3 for $\rho = 0.008$, Fig. 4 for $\rho = 0.004$, Fig. 5 for $\rho = 0.002$ and Fig. 6 for $\rho = 0.001$, respectively. It can be seen that the ML estimator is efficient, i.e., attains the CRB, as long as the SNR at each index is roughly

Fig. 1. Signal to noise ratio at different sample indices (due to the exponential decay). The dashed line at 0 dB is shown for reference.

> 8 dB. It is clear that for both LR and ML estimators, there is an optimal, in the MSE sense, length N which is determined by ρ (and of course the SNR). Worth noting is also that the minimas of LR and ML do not perfectly coincide (despite having the same bias), because of the lower ML estimation variance. From the four figures, a general rule-of-thumb for the ML and LR estimators can be formulated as to simply ensure that the SNR at each sample is positive with a margin of a couple dB. This is similar to what is proposed in [26] for the backward integration case. However, to implement this rule-of-thumb in practice is non-trivial because of the ρ dependence (and in the backward integration case also the parameter n_B , see section III-A).

As for MLN and hybrid MLN, it can be seen that the MSE of both methods is very similar, even though the hybrid method uses only up to 250 samples to estimate σ_v . Both methods attain the CRB up to approximately the same point as the ML estimator that does not take the noise into account, but instead of showing performance degradation, the MSE flattens out as N increases. Interesting to note is that the MSE of MLN and hybrid MLN flattens out at a lower MSE than the ML minimum. This means that using either MLN or hybrid MLN, not only eliminates the problem of selecting an appropriate N , but also makes it possible to achieve a lower estimation error. The simulated bias of the hybrid method is finally shown in Fig. 7 and it can be seen that it is practically unbiased.

Fig. 2. Theoretical and simulated bias of the LR estimator (upper plot) and ML estimator (lower plot). Full-drawn lines illustrate theoretical bias and the corresponding discrete symbols indicate simulated bias.

VII. A NOTE ON (BLIND) REVERBERATION TIME ESTIMATION ACCURACY

In blind reverberation time estimation, the reverberant signal is typically segmented into different sections corresponding to diffusive reverberation decay, and decay rate estimation is applied to *each segment*. One T_{60} estimate is then obtained by fusing the different decay rate estimates, see e.g., [12]. Thus, the accuracy of the T_{60} estimate does not only depend on the actual decay rate estimation, but also on the type of fusing used, the signal segmentation, as well as the correctness of the model. Nevertheless,

the following text briefly illustrates the T_{60} estimate accuracy given a single ρ estimate.

The case where $\rho = 0.002$, corresponding to a reverberation time T_{60} of 0.43 s for a sample rate F_s of 8 kHz, is first considered. In Fig. 4, for $N = 600$ (which is around the point where the noise starts affecting the MSE) the MSE of both ML and MLN is -76 dB, which implies standard deviation of around 1.67×10^{-4} , assuming no bias. If the error is normally distributed, about 95% of the estimates lie within ± 2 standard deviations, which in turn (using the relation in (3)) means that 95% of the reverberation time estimates lie between $T_{60} - 0.06$ s and $T_{60} + 0.09$ s. For the LR approach, the MSE when $N = 600$ is approximately -72 dB, meaning that the reverberation time estimate resides with 95% confidence between $T_{60} - 0.09$ s and $T_{60} + 0.15$ s.

In the case where $\rho = 0.008$, corresponding to a reverberation time T_{60} of 0.11 s for a sample rate F_s of 8 kHz, on the other hand, the MSE for ML and MLN is approximately -61 dB for $N = 200$. Hence, in this case, 95% of the estimates lie between $T_{60} - 0.02$ s and $T_{60} + 0.03$ s, and for the LR approach with a MSE of -57 dB, the 95% confidence is between $T_{60} - 0.03$ s and $T_{60} + 0.06$ s.

An interesting observation is that the performance of the “complete” blind reverberation time estimators in [20], operating on real speech signals, seem somewhat poor in comparison to the numbers presented above. (Although it should of course be noted that a straight-forward comparison is not sensible, due to the facts discussed in the beginning of this section.) This is most likely caused by imperfections of the reverberation decay detection/segmentation method(s), as well as model mismatch. The latter is supported by the fact that the performance indeed improves with whitening [19], and that assuming Laplacian/Exponential distribution instead of Gaussian in the model (see (1)) does not significantly affect the estimation performance [22], [23], and is an interesting subject for further research.

VIII. CONCLUSION

The basic methods for decay rate estimation used for blind reverberation time estimation originally presented in [7], [15] and [18] (here denoted LR, ML and MLN, respectively) have been analyzed and compared both with and without the influence of interfering noise under the assumption of i.i.d. Gaussian processes. It was shown that LR is unbiased if no smoothing is applied and that the estimation variance in the absence of noise is consistently about twice that of the ML estimator. Under the influence of interfering stationary noise, both LR and ML methods suffer from essentially the same level of performance degradation/bias. It was shown that in this case there exists an optimal, in the MSE sense, selection of the observation window length N , which depends on the SNR and decay factor ρ . A simple rule-of-thumb was formulated: ensure that the SNR at each sample is positive with a couple of dB margin.

Fig. 3. Theoretical and simulated MSE of the compared estimators for $\rho = 0.008$. Full-drawn lines illustrate theoretical MSE and the corresponding discrete symbols indicate simulated MSE.

Fig. 4. Theoretical and simulated MSE of the compared estimators for $\rho = 0.004$. Full-drawn lines illustrate theoretical MSE and the corresponding discrete symbols indicate simulated MSE.

The MSE of the ML approach is the same as MLN as long as the SNR is positive for all observed samples. As N is increased further, to also include samples with negative SNR, the MSE of the MLN is maintained while the MSE of the ML estimator deteriorates. Moreover, the ML has some bias, even in situations where the SNR is fairly high for all samples.

Fig. 5. Theoretical and simulated MSE of the compared estimators for $\rho = 0.002$. Full-drawn lines illustrate theoretical MSE and the corresponding discrete symbols indicate simulated MSE.

Fig. 6. Theoretical and simulated MSE of the compared estimators for $\rho = 0.001$. Full-drawn lines illustrate theoretical MSE and the corresponding discrete symbols indicate simulated MSE.

Finally, a hybrid approach was presented, comprising elements of both ML and MLN. The hybrid method showed similar performance to MLN, while requiring significantly less computational complexity. The method of choice in the considered environment is thus clearly the proposed method (or MLN if high computational complexity is no issue). In the case of negligible interfering additive noise level, ML

Fig. 7. Simulated bias of the hybrid ML estimator.

is most likely preferred over LR, owing to the lower estimation variance.

APPENDIX

DERIVATION OF (37)

First of all, it should be recognized that, using the substitution $\tau = -\exp(2\rho x + \nu)$, the relation

$$\int_{N_1}^{N_2} \text{Li}_s(-e^{2\rho x + \nu}) dx = \frac{1}{2\rho} \int_{N_3}^{N_4} \frac{\text{Li}_s(\tau)}{\tau} d\tau, \quad (55)$$

where $N_3 = -\exp(2\rho N_1 + \nu)$ and $N_4 = -\exp(2\rho N_2 + \nu)$ is obtained. Recognizing the repeated integral representation of the polylogarithm [31], gives that (55) can be written as

$$\frac{1}{2\rho} (\text{Li}_{s+1}(N_4) - \text{Li}_{s+1}(N_3)). \quad (56)$$

Now, since $\text{Li}_1(z) = -\ln(1 - z)$, using integration by parts yields that (36) can be re-written as

$$-\frac{6}{N(N^2 - 1)} \left(-\frac{N - 1}{4\rho} (\text{Li}_2(\beta_2) + \text{Li}_2(\beta_1)) + \frac{1}{4\rho^2} (-\text{Li}_3(\beta_2) + \text{Li}_3(\beta_1)) \right), \quad (57)$$

where $\beta_1 = -\exp(\nu + \rho(N - 1))$ and $\beta_2 = -\exp(\nu - \rho(N - 1)) = -\sigma_d^2/\sigma_v^2$. Finally, (57) is simplified to

$$\frac{3}{2} \left(\frac{\text{Li}_2(\beta_2) + \text{Li}_2(\beta_1)}{N(N + 1)\rho} + \frac{\text{Li}_3(\beta_2) - \text{Li}_3(\beta_1)}{N(N^2 - 1)\rho^2} \right). \quad (58)$$

REFERENCES

- [1] P. A. Naylor and N. D. Gaubitch, *Speech dereverberation*. Springer, 2010.
- [2] J. Benesty, J. Chen, and Y. Huang, *Microphone array signal processing*. Springer, 2008, vol. 1.
- [3] ISO 3382, *Acoustics - Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters*, 2nd edition, International Organization for Standardization, Geneva, Switzerland, 1997.
- [4] M. Kuster, "Reliability of estimating the room volume from a single room impulse response," *The Journal of the Acoustical Society of America*, vol. 124, pp. 982–993, 2008.
- [5] E. A. P. Habets, "Multi-channel speech dereverberation based on a statistical model of late reverberation," in *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Sig. Proc. (ICASSP)*, March 2005.
- [6] H. W. Löllman and P. Vary, "Low delay noise reduction and dereverberation for hearing aids," *EURASIP Journal on advances in sig. proc.*, Jan. 2001.
- [7] K. Lebart, J. M. Boucher, and P. N. Denbigh, "A new method based on spectral subtraction for speech dereverberation," *Acta Acustica united with Acustica*, vol. 87, no. 3, pp. 359–366, May/June 2001.
- [8] E. Larsen, C. D. Schmitz, C. R. Lansing, W. D. O'Brien, B. C. Wheeler, and A. S. Feng, "Acoustic scene analysis using estimated impulse responses," in *Proc. of the 37th Asilomar Conf. on Sig., Sys. and Comp.*, vol. 1, Nov. 2003, pp. 725–729.
- [9] J. Y. Wen, E. A. Habets, and P. A. Naylor, "Blind estimation of reverberation time based on the distribution of signal decay rates," in *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Sig. Proc. (ICASSP)*, March 2008, pp. 329–332.
- [10] T. J. Cox, F. Li, and P. Darlington, "Extracting room reverberation time from speech using artificial neural networks," *Journal of the audio engineering society*, vol. 49, no. 4, pp. 219–230, 2001.
- [11] R. Talmon and E. A. P. Habets, "Blind reverberation time estimation by intrinsic modeling of reverberant speech," in *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Sig. Proc. (ICASSP)*. IEEE, 2013, pp. 156–160.
- [12] J. Eaton, N. D. Gaubitch, and P. A. Naylor, "Noise-robust reverberation time estimation using spectral decay distributions with reduced computational cost," in *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Sig. Proc. (ICASSP)*, May 2013.
- [13] T. Falk, C. Zheng, and W. Y. Cha, "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech," *IEEE Trans. Audio, Speech, Lang. Proc.*, vol. 18, no. 7, pp. 1766–1774, Sept. 2010.
- [14] T. d. M. Prego, A. A. de Lima, S. L. Netto, B. Lee, A. Said, R. W. Schafer, and T. Kalker, "A blind algorithm for reverberation-time estimation using subband decomposition of speech signals," *The Journal of the Acoustical Society of America*, vol. 131, pp. 2811–2816, 2012.
- [15] R. Ratnam, D. L. Jones, B. C. Wheeler, W. D. O'Brien, C. R. Lansing, and A. S. Feng, "Blind estimation of reverberation time," *Journal of the Acoustical Society of America*, vol. 114, no. 5, pp. 2877–2892, Nov. 2003.
- [16] R. Ratnam, D. L. Jones, and W. D. O'Brien, "Fast algorithms for blind estimation of reverberation time," *IEEE Sig. Proc. Letters*, vol. 11, no. 6, pp. 537–540, June 2004.
- [17] H. W. Löllmann, E. Yilmaz, M. Jeub, and P. Vary, "An improved algorithm for blind reverberation time estimation," in *Proc. of Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Aug. 2010.
- [18] H. W. Löllmann and P. Vary, "Estimation of the reverberation time in noisy environments," in *Proc. Intl. Workshop Acoustic Echo and Noise Control (IWAENC)*, vol. 4, Sept. 2008.
- [19] A. Keshavarz, S. Mosayyebpour, M. Biguesh, T. A. Gulliver, and M. Esmaeili, "Speech-model based accurate blind reverberation time estimation using an LPC filter," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1884–1893, 2012.

- [20] N. D. Gaubitch, H. W. Löllmann, M. Jeub, T. H. Falk, P. A. Naylor, P. Vary, and M. Brookes, "Performance comparison of algorithms for blind reverberation time estimation from speech," in *Proc. of Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Sept. 2012.
- [21] H. Kuttruff, *Room Acoustics*, 4th ed. Taylor & Francis, Oct. 2000.
- [22] J. Tariqullah and W. Wang, "Blind reverberation time estimation based on Laplace distribution," in *Proc. of the 20th European Sig. Proc. Conf. (EUSIPCO)*, Aug. 2012, pp. 2050–2054.
- [23] C. Schüldt and P. Händel, "Blind low-complexity estimation of reverberation time," in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2013.
- [24] N. López, Y. Grenier, and I. Bourmeyster, "Low variance blind estimation of the reverberation time," in *Proc. of Intl. Workshop on Acoustic Signal Enhancement (IWAENC)*, 2012.
- [25] S. M. Kay, "Fundamentals of statistical signal processing, volume I: Estimation theory," 1993.
- [26] D. R. Morgan, "A parametric error analysis of the backward integration method for reverberation time estimation," *The Journal of the Acoustical Society of America*, vol. 101, pp. 2686–2693, 1997.
- [27] S. Vesa and A. Harma, "Automatic estimation of reverberation time from binaural signals," in *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Sig. Proc. (ICASSP)*, vol. 3. IEEE, 2005, pp. 281–284.
- [28] M. R. Schroeder, "New method of measuring reverberation time," *Journal of the Acoustical Society of America*, vol. 37, no. 3, pp. 409–412, 1965.
- [29] N. Xiang, "Evaluation of reverberation times using a nonlinear regression approach," *The Journal of the Acoustical Society of America*, vol. 98, pp. 2112–2121, 1995.
- [30] M. Abramowitz and I. A. Stegun, "Handbook of mathematical function with formulas, graphs, and mathematical tables," *National Bureau of Standards, Applied Mathematics Series*, vol. 55, 1970.
- [31] D. Wood, "The computation of polylogarithms," University of Kent, Computing Laboratory, University of Kent, Canterbury, UK, Tech. Rep. 15-92*, June 1992. [Online]. Available: <http://www.cs.kent.ac.uk/pubs/1992/110>
- [32] E. Hänsler and G. Schmidt, *Acoustic echo and noise control: a practical approach*. Wiley, 2004.
- [33] Y.-X. Yao and S. M. Pandit, "Cramér-rao lower bounds for a damped sinusoidal process," *IEEE Transactions on Signal Processing*, vol. 43, no. 4, pp. 878–885, 1995.

Christian Schüldt (S'06-M'13) was born in Stockholm, Sweden. He received the B.Sc. and M.Sc. degrees in electrical engineering from the Royal Institute of Technology KTH, Stockholm, Sweden in 2001 and 2004, respectively, and the Ph.D. degree in Telecommunication from Blekinge Institute of Technology, Karlskrona, Sweden, in 2012. Currently, he is CTO of Limes Audio AB, Umeå, Sweden, as well as a part-time post-doc researcher at the Royal Institute of Technology KTH, Stockholm, Sweden. His areas of professional interest include low-complexity methods for speech enhancement and audio signal processing

in general.

Peter Händel (S'88-M'94-SM'98) received the Ph.D. degree from Uppsala University, Uppsala, Sweden, in 1993. From 1987 to 1993, he was with Uppsala University. From 1993 to 1997, he was with Ericsson AB, Kista, Sweden. From 1996 to 1997, he was also with the Tampere University of Technology, Tampere, Finland. Since 1997, he has been with the Royal Institute of Technology, Stockholm, Sweden, where he is currently a Professor of Signal Processing. From 2000 to 2006, he worked part time with the Swedish Defence Research Agency. In 2010, he spent some time at the Indian Institute of Science, Bangalore,

India, as a Guest Professor. He is currently a Guest Professor with the University of Gävle, Gävle, Sweden.

Dr. Händel has served as an Associate Editor for the IEEE Transactions on Signal Processing.