An Improved Algorithm for Blind Reverberation Time Estimation

Heinrich W. Löllmann, Emre Yilmaz, Marco Jeub, and Peter Vary
Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, 52056 Aachen, Germany
{loellmann,yilmaz,jeub,vary}@ind.rwth-aachen.de

***Abstract*—An improved algorithm for the estimation of the reverberation time (RT) from reverberant speech signals is presented. This blind estimation of the RT is based on a simple statistical model for the sound decay such that the RT can be estimated by means of a maximum-likelihood (ML) estimator. The proposed algorithm has a significantly lower computational complexity than previous ML-based algorithms for RT estimation. This is achieved by a downsampling operation and a simple pre-selection of possible sound decays. The new algorithm is more suitable to track time-varying RTs than related approaches. In addition, it can also estimate the RT in the presence of (moderate) background noise.**

The proposed algorithm can be employed to measure the RT of rooms from sound recordings without using a dedicated measurement setup. Another possible application is its use within speech dereverberation systems for hands-free devices or digital hearing aids.

***Index Terms*—reverberation time, blind estimation, low complexity, speech dereverberation**

I. Introduction

The *reverberation time* (RT) is an important quantity for the characterization of enclosed auditory spaces [1]. This measure is commonly used to assess the amount of room reverberation or its effects. The RT is defined as the time interval in which the energy of a steady-state sound field decays 60 dB below its initial level after switching off the excitation source. The measurement of the RT out of a sound decay can be done by the interrupted noise method [2]. Schroeder has proposed a method to obtain the RT directly from a measured *room impulse response* (RIR) instead of calculating it by the ensemble average of different sound decays [3].

However, such methods can not always be applied for *reverberation time estimation* (RTE). In many cases, it is necessary to perform a *blind* RTE, i.e., the RT has to be determined from a reverberant signal without knowing the excitation signal or room geometry. One example is the measurement of the RT within occupied rooms where the use of noisy excitation signals can disturb or irritate the occupants. Another prominent example are speech enhancement systems where knowledge about the RT can be exploited to perform speech dereverberation [4], [5].

A semi-blind estimation of the RT is presented in [6] where the room characteristics are 'learned' by using a neural network approach. In the context of acoustic echo cancelation, an estimate of the impulse response between loudspeaker and microphone is available from which the RT can be determined [7], [8].

A blind estimation of short RTs ($T_{60} < 0.6\,\mathrm{s}$), which is based on an estimate of the pitch period of speech signals, is proposed in [9]. An approach for a blind RTE in the frequency-domain is presented in [10]. The distributions of speech decays are thereby measured in the short-term DFT domain where the change of these distributions due to reverberation is exploited to estimate the RT. A training is required to calibrate the needed 'mapping parameters'.

An alternative method, which does not require such a calibration, is proposed in [11], [12]. The RT is estimated in the time-domain and relies on a *maximum-likelihood estimation* (MLE). In contrast to the blind RTE of [4], a sound decay detection is not employed, which results in a more accurate and robust estimation. In [13], it is shown how the ML-based RTE of [11] can be extended to estimate the RT in the presence of background noise, which is of interest, e.g., for speech enhancement systems [5], [14].

A problem of these blind methods for RTE is their rather high computational complexity and their slow convergence towards changing RTs. These problems are addressed in this contribution. The devised algorithm for blind RTE is also based on a MLE, but has a significantly lower computational complexity than previous proposals. In addition, the new algorithm accounts also for time-varying RTs.

The paper is organized as follows: In Sec. II, the underlying sound decay model and maximum-likelihood (ML) estimation of the RT are introduced. The new algorithm is described in Sec. III where simulation examples are provided by Sec. IV. The main results of this contribution are summarized in Sec. V.

II. Sound Decay Model & ML Estimation

A reverberant speech signal is considered which is given by a speech signal $s(k)$ convolved with a time-varying RIR $h(\eta, k)$ of (possibly infinite) length L_h:

$$z(k) = \sum_{\eta=0}^{L_h-1} s(k-\eta) \cdot h(\eta, k) \tag{1}$$

with k marking the discrete time index. Within a speech pause

$$s(k-\eta) \begin{cases} \approx 0 & \text{for} \quad \eta = 0, 1, \ldots, L_\mathrm{o} - 1 \\ \neq 0 & \text{for} \quad \eta = L_\mathrm{o}, \ldots, L_h - 1, \end{cases} \tag{2}$$

the room reverberation causes a sound decay $d(k)$ since

$$z(k) = \underbrace{\sum_{\eta=0}^{L_\mathrm{o}-1} s(k-\eta)\cdot h(\eta,k)}_{\approx\, 0} + \underbrace{\sum_{\eta=L_\mathrm{o}}^{L_h-1} s(k-\eta)\cdot h(\eta,k)}_{\doteq\, d(k)} \quad (3)$$

where it is assumed that $h(L_\mathrm{o},k) \neq 0$. The sound decay $d(k)$ is *modeled* by a discrete random process

$$d_\mathrm{m}(k) = A_\mathrm{r}\, v(k)\, e^{-\rho\, k\, T_\mathrm{s}}\, \epsilon(k) \quad (4)$$

with real amplitude $A_\mathrm{r} > 0$, decay rate ρ and $\epsilon(k)$ marking the unit step sequence. The variable $T_\mathrm{s} = 1/f_\mathrm{s}$ marks the sampling period and $v(k)$ is a sequence of i.i.d. random variables with zero mean, variance of one and normal distribution $\mathcal{N}(0,1)$. The energy decay curve for the corresponding time-continuous sound decay model reads

$$E_{\tilde{d}}(t) \doteq E\left\{\tilde{d}_\mathrm{m}^2(t)\right\} = A_\mathrm{r}^2\, e^{-2\,\rho\, t}\,\tilde{\epsilon}(t) \quad (5)$$

where the tilde indicates the time-continuous counterparts to the discrete quantities of Eq. (4). A relation between the *decay rate* ρ and the *reverberation time* T_{60} can be established by the requirement

$$10\,\log_{10}\left(\frac{E_{\tilde{d}}(0)}{E_{\tilde{d}}(T_{60})}\right) \overset{!}{=} 60 \quad (6)$$

such that

$$T_{60} = \frac{3}{\rho\,\log_{10}(e)} \approx \frac{6.908}{\rho}\ . \quad (7)$$

Due to this relation, the terms decay rate and RT will be used interchangeably in the following.

According to the above model, the decay $d(k)$ is represented by a random variable with Gaussian probability density function (PDF)

$$p_{d(k)}(x) = \frac{1}{\sqrt{2\,\pi}\,\xi(k)}\,\exp\left\{-\frac{x^2}{2\,\xi^2(k)}\right\} \quad (8)$$

where

$$\xi(k) = A_\mathrm{r}\, a^k\, \epsilon(k) \quad \text{and} \quad a = e^{-T_\mathrm{s}\,\rho}\ . \quad (9)$$

The sequence $d(k)$ for $k \in \{0,\ldots,N-1\}$ is modeled by N independent random variables with zero mean and non-identical PDFs having normal distributions. This allows to derive a *maximum-likelihood* (ML) estimator for the unknown decay rate or RT, respectively, [11], [13]. The decay rate ρ is estimated from a given sound decay $d(k)$ by finding the maximum

$$\hat{\rho}^{(\mathrm{ML})} = \max_{\rho}\{\mathcal{L}(\rho)\} \quad (10a)$$

of the log-likelihood function

$$\mathcal{L}(\rho) = -\frac{N}{2}\left((N-1)\ln(a) + \ln\left(\frac{2\,\pi}{N}\sum_{i=0}^{N-1} a^{-2\,i}\,d^2(i)\right) + 1\right). \quad (10b)$$

The ML estimate for the RT $\widehat{T}_{60}^{(\mathrm{ML})}$ is obtained by Eq. (7).

Eq. (4) can also be seen as a simple statistical model for the RIR, which considers only the effects of late reflections and models them as diffuse noise. Accordingly, the MLE can also be used to estimate the RT out of a measured RIR. The model of Eq. (4) is rather course and different generalizations of the original ML-based RTE of [11] have been proposed. In [13], a MLE of the RT in the presence of additive noise is derived. A MLE which accounts for multiple decay rates (associated with early and late room reflections) is proposed in [15]. However, such a model is not considered here as it causes a significantly increased computational complexity.

III. Efficient RT Estimation

For the blind RTE, the reverberant speech signal $z(k)$ is first downsampled by R sample instants

$$x(n) = z(n\,R), \quad R \in \mathbb{N} \quad (11)$$

with n marking the discrete time index after subsampling. This downsampling allows to reduce the computational complexity of the algorithm where the choice for R depends on the sampling frequency f_s. This approach is reasoned by the fact that the estimation of the RT by a MLE (or the Schroeder method) relies on an energy decay, cf., Eq. (5) and (6). The characteristic of this energy decay is also preserved if a (moderate) subsampling is applied.

The downsampled sequence is processed within frames of M samples shifted by M_Δ sample instants

$$x_\mathrm{f}(\lambda,m) = x(\lambda\,M_\Delta + m) \quad \text{with} \quad m = 0,1,\ldots,M-1 \quad (12)$$

and $\lambda \in \mathbb{N}$. In a first step, a pre-selection is conducted to detect possible sound decays, cf., Eq. (3). For this, the current frame $x_\mathrm{f}(\lambda,m)$ is divided into $L = M/P \in \mathbb{N}$ sub-frames

$$y(\lambda,l,\kappa) = x_\mathrm{f}(\lambda, l\,P + \kappa) \quad (13)$$

with $\kappa = 0,1,\ldots P-1$ and sub-frame index $l = 0,1,\ldots L-1$. Then it is checked whether the energy, maximum and minimum value of a sub-frame deviates from the successive sub-frame according to

$$\sum_{\kappa=0}^{P-1} y^2(\lambda,l,\kappa) > w_{l+1}^{(\mathrm{var})}\cdot\sum_{\kappa=0}^{P-1} y^2(\lambda,l+1,\kappa) \quad (14a)$$

$$\max_{\kappa}\{y(\lambda,l,\kappa)\} > w_{l+1}^{(\mathrm{max})}\cdot\max_{\kappa}\{y(\lambda,l+1,\kappa)\} \quad (14b)$$

$$\min_{\kappa}\{y(\lambda,l,\kappa)\} < w_{l+1}^{(\mathrm{min})}\cdot\min_{\kappa}\{y(\lambda,l+1,\kappa)\} \quad (14c)$$

with sub-frame counter $l = 0,1,\ldots L-2$ and weighting factors $0 \le w_l \le 1$. If one of these conditions is violated, it is checked whether the counter l has reached a minimum value $1 < l_\mathrm{min} < L-2$. If this is not the case, the comparison is aborted and the next signal frame $x_\mathrm{f}(\lambda+1,m)$ is processed. Otherwise, the consecutive sub-frames for which Eq. (14) applies are detected as a possible sound decay. For this segment, the RT is calculated by means of Eq. (10) for a finite set of RT values (decay rates).

A new ML estimate is used to update a histogram comprising the last K_f ML estimates for the RT. The RT $\widehat{T}_{60}^{(1)}(\lambda)$

associated with the maximum of the histogram is taken as current RT estimates. (The maximum instead of the first peak can be taken as this histogram contains due to the pre-selection no significant number of outlier as for the approaches of [11], [13].) The variance for the estimated RT is reduced by a recursive smoothing such that the final estimate is given by

$$\widehat{T}_{60}(\lambda) = \beta(\lambda) \cdot \widehat{T}_{60}(\lambda - 1) + \big(1 - \beta(\lambda)\big) \cdot \widehat{T}_{60}^{(1)}(\lambda) \qquad (15)$$

with $0 < \beta(\lambda) < 1$.

The choice for the time-varying smoothing factor $\beta(\lambda)$ as well as K_f is subject to a trade-off: High values reduce the variance for the RT estimate, but changes are detected more slowly and vice versa. In order to alleviate this problem with low complexity, a second histogram is introduced which is determined by the last $K_\mathrm{s} < K_\mathrm{f}$ ML estimates. If the RT estimates obtained from this second histogram $\widehat{T}_{60}^{(2)}(\lambda)$ deviates from that of the first one $\widehat{T}_{60}^{(1)}(\lambda)$ for a certain period

$$\left|\widehat{T}_{60}^{(1)}(\lambda) - \widehat{T}_{60}^{(2)}(\lambda)\right| > \epsilon_T \quad \text{for} \quad \lambda = \lambda_1, \ldots, \lambda_Q, \qquad (16)$$

the estimate of the second histogram $\widehat{T}_{60}^{(2)}(\lambda)$ is used for Eq. (15) and the first histogram is filled with the values of the second one. In this case, a low smoothing factor $\beta(\lambda) < 0.5$ is used for Eq. (15) where a high smoothing factor $\beta(\lambda) > 0.9$ is taken otherwise.

Some properties of the new approach should be noticed. The detection of a possible sound decay according to Eq. (14) reduces the high computational burden of executing Eq. (10) within fixed time intervals as done in [11]–[13].[1] Another benefit (and difference) is that the frame length N for Eq. (10) is now adapted to the length of the detected sound decay and not constant. This enables to estimate a higher range of RTs (as demonstrated in the next section). In contrast to the RTE of [4], a false detection for a sound decay does not directly lead to an overestimated RT as the final estimate is obtained from a histogram which results in a more robust estimate.

IV. Simulation Examples

The performance of the new algorithm for RTE shall be illustrated by some simulation examples. For this purpose, a speech signal is convolved with a RIR according to Eq. (1) for a sampling frequency of $f_\mathrm{s} = 16\,\mathrm{kHz}$. The coefficients of the RIR are switched two times at instant k_0 and k_1 to analyze the tracking of a changing RT. The employed RIRs are taken from the AIR database [16] (and downsampled to 16 kHz). The first RIR with a RT of 0.25 s is measured in a low reverberant studio booth at a distance of 50 cm between loudspeaker and microphone. The second RIR with a RT of 0.67 s is measured in a reverberant office room with a loudspeaker-microphone distance of 300 cm. Both RIRs are measured without a dummy head and the calculation of the actual RTs is based on the Schroeder method as described in [13].

The main parameters that are used for the improved RTE are listed in Table I. The evaluation of Eq. (10) is performed for discrete decay rates corresponding to $T_{60} \in \{0.1\,\mathrm{s}, 0.105\,\mathrm{s}, \ldots, 1.5\,\mathrm{s}\}$ and a bin size of $0.05\,\mathrm{s}$ is taken for the histograms.

TABLE I
Main parameters of the new algorithm for RTE ($f_\mathrm{s} = 16\,\mathrm{kHz}$).

R	M	M_Δ	L	l_min	$\beta(\lambda)$	ϵ_T	K_s	K_f	Q
5	128	25	7	3	{0.995,0.2}	0.2 s	400	20	30

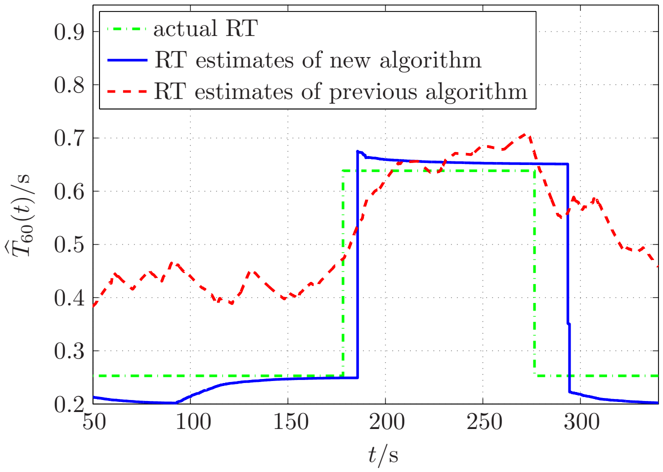

Fig. 1. Comparison of the new algorithm for blind RTE with the algorithm of [13] for a time-varying RIR.

The new algorithm is compared with the blind RTE of [13].[2] The RTs estimated over time by the two algorithms are plotted in Fig. 1. It can be observed that the proposed RTE provides a more accurate estimate of the RT than the previous approach [13]: The deviation from the actual RT is much lower and the change of the RT is detected more accurately. (The detection of time-varying RTs can be improved by a different parameter setting or smoothing for the reference algorithm [13], but this in turn increases the variance for the RT estimate.) The overestimation of the reference algorithm [13] for low RTs is attributed to the fixed buffer length used for the ML estimation. As a consequence, this buffer contains not only the sound decay, but also the following tail which causes the overestimation. The use of a smaller buffer length can alleviate this problem but decreases the estimation accuracy for higher RTs. The new algorithm solves this problem by adapting the buffer length to that of the detected sound decay.

The new RT estimation possesses also a significantly lower computational complexity; the execution time of its MATLAB implementation was more than 3 times faster than for the reference algorithm.

A blind RTE has often to be performed in noisy environments. An example is the use of a blind RTE for speech dereverberation and noise reduction, e.g., [5], [14]. A noise reduction can be applied to the noisy and reverberant speech in a pre-processing step which, however, can only achieve a partial noise reduction. Hence, the RTE has to cope (at least) with some residual noise. Therefore, the RT estimation out

[1] The 'fast online algorithm' of [12] updates the log-likelihood function efficiently for each sample instant k. Here, the ML estimation is not calculated at each sample instant, but only if a sound decay is detected which results in a much lower computational load.

[2] A comparison with the blind RTE of [11], [12] requires an adaptation of the algorithm to account for time-varying RTs as in [13].

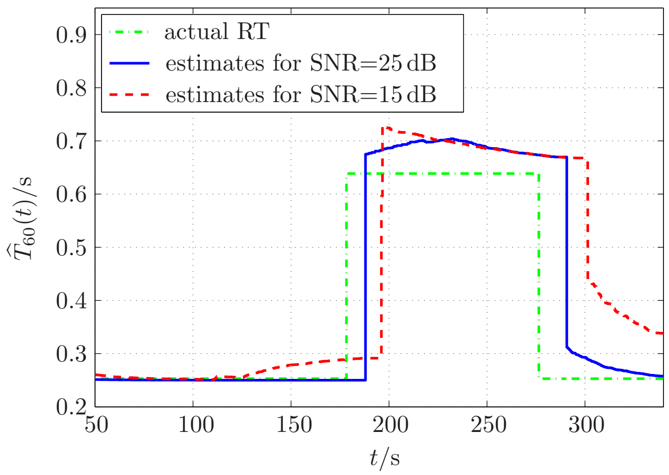

Fig. 2. Evaluation of the new RTE for a time-varying RIR where the reverberant speech signal is disturbed by additive, white Gaussian noise with different SNRs. (The corresponding curve for an RT estimation without noise is given by Fig. 1.)

of a reverberant and noisy speech signal is now considered. The reverberant speech signal is distorted by additive, white Gaussian noise with different signal-to-noise-ratios (SNRs). The obtained results are shown in Fig. 2. A comparison with Fig. 1 reveals that the noise leads to an overestimation of the RT, especially for lower SNRs. Another consequence is that a changing RT is detected more slowly as more speech decays are now deemed as unsuitable due to the noise. However, the presented RTE still provides feasible results for moderate noise. The noise causes that the sound decay immerges into a ‘noise floor’. The pre-selection procedure detects the actual sound decay and avoids to some extend that the noise floor is included in the segment used for the ML estimation.

In addition, a pre-denoising can be applied in case of a low SNR by means of a noise reduction system, cf., [13], [14]. The investigation of the improved RTE w.r.t. its application to speech enhancement is a topic of ongoing research and beyond the scope of this contribution.

V. Conclusions

An improved algorithm for the blind estimation of the RT is devised. It is based on a simple statistical model for the sound decay such that the RT can be estimated by a ML estimation. In contrast to previous approaches [11]–[13], the new algorithm exhibits a significantly lower computational complexity. This is achieved by a downsampling operation and an efficient pre-selection of possible sound decays. In addition, the new algorithm can track time-varying RTs with a much higher accuracy than related algorithms for blind RTE. The proposed method is also capable of estimating the RT in the presence of moderate background noise.

The presented algorithm for a blind RT estimation with low complexity is of interest, among others, for speech dereverberation in digital hearing aids [14], [17].

References

[1] H. Kuttruff, *Room Acoustics*, Taylor & Francis, London, 4th edition, 2000.

[2] ISO-3382, “Acoustics-Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters,” International Organization for Standardization, Geneva, Switzerland, 1997.

[3] M. R. Schroeder, “New Method of Measuring Reverberation Time,” *Journal of the Acoustical Society of America*, vol. 37, pp. 409–412, 1965.

[4] K. Lebart, J. M. Boucher, and P. N. Denbigh, “A New Method Based on Spectral Subtraction for Speech Dereverberation,” *acta acoustica - ACOUSTICA*, vol. 87, no. 3, pp. 359–366, 2001.

[5] E. A. P. Habets, *Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement*, Ph.D. thesis, Eindhoven University, Eindhoven, The Netherlands, June 2007.

[6] T. J. Cox, F. Li, and P. Dalington, “Extracting Room Reverberation Time From Speech Using Artificial Neural Networks,” *Journal of the Acoustical Society of America*, vol. 49, no. 4, pp. 219–230, 2001.

[7] M. Buck and A. Wolf, “Model-Based Dereverberation of Single-Channel Speech Signals,” in *Proc. of German Annual Conference on Acoustics (DAGA)*, Dresden, Germany, Mar. 2008, pp. 261–262.

[8] E. A. P. Habets, S. Gannot, and I. Cohen, “Dereverberation and Residual Echo Suppression in Noisy Environments,” in *Speech and Audio Processing in Adverse Environments*, E. Hänsler and G. Schmidt, Eds., chapter 6, pp. 185–227. Springer, Berlin, 2008.

[9] M. Wu and D. Wang, “A Pitch-Based Method for the Estimation of Short Reverberation Time,” *Acta Acustica United With Acustica*, vol. 92, pp. 337–339, 2006.

[10] J. Y. C. Wen, E. A. P. Habets, and P. A. Naylor, “Blind Estimation of Reverberation Time Based on the Distribution of Signal Decay Rates,” in *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Las Vegas (Nevada), USA, Apr. 2008, pp. 329–332.

[11] 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.

[12] R. Ratnam, D. L. Jones, and W. D. O’Brien, “Fast Algorithms for Blind Estimation of Reverberation Time,” *IEEE Signal Processing Letters*, vol. 11, no. 6, pp. 537–540, June 2004.

[13] H. W. Löllmann and P. Vary, “Estimation of the Reverberation Time in Noisy Environments,” in *Proc. of Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle (Washington), USA, Sept. 2008.

[14] H. W. Löllmann and P. Vary, “Low Delay Noise Reduction and Dereverberation for Hearing Aids,” *EURASIP Journal on Applied Signal Processing, Special Issue on Digital Signal Processing for Hearing Instruments*, vol. 2009, pp. 1–9, Jan. 2009.

[15] P. Kendrick, F. F. Li, and T. J. Cox, “Blind Estimation of Reverberation Parameters for Non-Diffuse Rooms,” *Journal of the Acoustical Society of America*, vol. 93, pp. 760–770, 2007.

[16] M. Jeub, M. Schäfer, and P. Vary, “A Binaural Room Impulse Response Database for the Evaluation of Dereverberation Algorithms,” in *Proc of. International Conference on Digital Signal Processing (DSP)*, Santorini, Greece, July 2009.

[17] M. Jeub, H. W. Löllmann, and P. Vary, “Blind Dereverberation for Hearing Aids with Binaural Link,” in *Proc. of ITG Conference on Speech Communication*, Bochum, Germany, Oct. 2010.