

Reverberation Time Estimation for Speech Processing Applications

Heinrich W. Löllmann and Peter Vary*

Institute of Communication Systems and Data Processing (ivd)

RWTH Aachen University, 52056 Aachen, Germany

e-mail: {loellmann, vary}@ind.rwth-aachen.de

Abstract

The reverberation time (RT) is an important measure for the characterization of reverberant environments, which can be determined by different acoustical measurement methods. However, they use mostly a special measurement setup and dedicated excitation signals, which is inapplicable for most speech processing algorithms requiring knowledge about the RT.

In this contribution, a new method is devised which allows to estimate the RT from noisy observations. It is based on a maximum likelihood (ML) estimation which is derived from a statistical model of the sound decay in reverberant and noisy enclosures. This allows to determine the RT from a measured sound decay (or room impulse response) despite the presence of noise. It is also shown, how the ML estimator can be used to determine the RT blindly from a noisy and reverberant speech signal. This blind RT estimation can be employed for the enhancement of noisy and reverberant speech signals.

Introduction

The *reverberation time* (RT) is a well-known and important measure for the characterization of reverberant enclosures. It is defined as the time span in which the energy of a steady-state sound field decays 60 dB below its initial level after switching-off the excitation source [1]. Knowledge about the RT is of interest, among others, for the characterization of acoustic environments [1], predicting the subjective preference of reverberant speech [2], or for the enhancement of distorted speech signals, e.g., [3, 4]. Accordingly, methods for *reverberation time estimation* (RTE) are a subject of interest for acousticians and engineers alike.

The RT can be determined by measuring the sound decay after turning-off the excitation source, e.g., by means of the interrupted noise method [5]. Schroeder has developed a method to calculate the ensemble average of different decay curves from the measured *room impulse response* (RIR) [6]. The Schroeder (impulse) method forms the basis for other approaches to estimate the RT by means of excitation sound sources, e.g., [7, 8]. The method of Xiang [8] allows also to estimate the RT from a noisy sound decay by means of non-linear regression. The RT is calculated by an iterative procedure which, however, relies very much on a good initial guess for the first iteration and does not necessarily converge.

For speech enhancement systems, the use of an excitation sound source to acquire the RT is of course impractical. Instead, the RT must be estimated *blindly* from a reverberant and mostly noisy speech signal. Methods for a (semi-)blind RT estimation have been proposed, e.g., in [9, 3, 10]. In [9], room characteristics are 'learned' by using a neural network approach. Other methods try to detect speech offsets (gaps) in the speech signal to measure the sound decay using either one or two microphones [3, 10]. Algorithms for an entirely blind estimation of the RT are presented in [11, 12, 13]. However, all these proposals for a (partly) blind estimation of the RT deal not (explicitly) with the impairments due to additive noise.

Hence, estimating the RT in *noisy* environments and with little efforts is still a challenging problem. In this contribution, we will review recent proposals to tackle this problem by means of a *maximum likelihood* (ML) estimator and outline its application for speech enhancement based on [14, 15].

Sound Decay Model

It is assumed that the observed sequence $y(k)$ contains the sound decay due to reverberation $h_M(k)$ and additive noise $n(k)$:

$$y(k) = h_M(k) + n(k). \quad (1)$$

The noise sequence $n(k)$ is assumed to be uncorrelated with $h_M(k)$ and represents i.i.d. random variables with zero mean and normal distribution $\mathcal{N}(0, \sigma_n^2)$. The sound decay is modeled as a discrete random process

$$h_M(k) = A_r v(k) e^{-\rho k T_s} \epsilon(k) \quad (2)$$

with real amplitude $A_r > 0$. The variable k marks the discrete sample index and $\epsilon(k)$ the unit step sequence. The parameter $T_s = 1/f_s$ represents the sampling period and $v(k)$ is a sequence of i.i.d. random variables with zero mean and normal distribution $\mathcal{N}(0, 1)$. Eq. (2) 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. The energy decay curve for the corresponding time-continuous sound decay model reads

$$E_{\tilde{h}}(t) \doteq E \left\{ \tilde{h}_M^2(t) \right\} = A_r^2 e^{-2\rho t} \tilde{\epsilon}(t) \quad (3)$$

where the tilde indicates the time-continuous counterparts to the discrete quantities of Eq. (2). A relation between the *decay rate* ρ and the *reverberation time* T_{60}

*This work was supported by GN ReSound, Eindhoven, The Netherlands.

can be established by the requirement

$$10 \log_{10} \left(\frac{E_{\tilde{h}}(0)}{E_{\tilde{h}}(T_{60})} \right) \stackrel{!}{=} 60 \quad (4)$$

which leads to the equation

$$T_{60} = \frac{3}{\rho \log_{10}(e)} \approx \frac{6.908}{\rho}. \quad (5)$$

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

According to our model, $y(k)$ is a random variable with the Gaussian probability density function (PDF)

$$p_{y(k)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}\xi(k)} \exp \left\{ -\frac{x^2}{2\sigma^2\xi^2(k)} \right\} \quad (6)$$

$$\text{with } \xi(k) = \sqrt{A_r^2 \cdot a^{2k} \cdot \epsilon(k) + \sigma_n^2} \text{ and } a = e^{-T_s \rho}. \quad (7)$$

Hence, the sequence $y(k)$ for $k \in \{0, \dots, N-1\}$ consists of N independent random variables with zero mean and *non-identical* PDFs having normal distributions $\mathcal{N}(0, \xi^2(k) \cdot \sigma^2)$.

Maximum Likelihood Estimation

The model introduced before enables the use of a maximum likelihood (ML) estimator (cf., [16]) for the RT. The likelihood function (joint PDF) for an observed sequence of N (noisy) samples $y(k)$ is derived from Eq. (6):

$$L_f(y, \xi, \sigma) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}} \prod_{i=0}^{N-1} \xi(i)} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=0}^{N-1} \frac{y^2(i)}{\xi^2(i)} \right\} \quad (8)$$

which yields the following *log-likelihood function* (LLF)

$$\begin{aligned} \mathcal{L}(y, \xi, \sigma) &= \ln(L_f(y, \xi, \sigma)) \\ &= -\frac{N}{2} \ln(2\pi\sigma^2) - \sum_{i=0}^{N-1} \ln(\xi(i)) \\ &\quad - \frac{1}{2\sigma^2} \sum_{i=0}^{N-1} \frac{y^2(i)}{\xi^2(i)} \end{aligned} \quad (9)$$

with $\ln(\cdot)$ representing the natural logarithm. The unknown *damping factor* a (and thus T_{60}) can be estimated by the maximum of the LLF

$$\hat{a} = \arg \left\{ \underset{A_r, a, \sigma, \sigma_n}{\text{maximum}} \{ \mathcal{L}(y, \xi, \sigma) \} \right\} \quad (10)$$

where the dependence from the variables (y, ξ, σ) will be omitted in the following to simplify the notation. The noise variance σ_n^2 can be assumed to be known as it can be determined by the noise floor following the sound decay. Eq. (10) can be solved by setting the partial derivatives towards the unknowns equal to zero which leads after some manipulations to the new LLF [14]

$$\mathcal{L} = -\frac{N}{2} \left(\ln \left(\frac{2\pi}{N} \sum_{i=0}^{N-1} \frac{y^2(i)}{\xi^2(i)} \right) + 1 \right) - \sum_{i=0}^{N-1} \ln(\xi(i)). \quad (11)$$

The unknown damping factor is then determined by

$$\hat{a} = \arg \left\{ \underset{A_r, a}{\text{maximum}} \{ \mathcal{L} \} \right\}. \quad (12)$$

This approach is termed as *generalized maximum likelihood* (GML) estimator as it enables a RTE in noisy environments in contrast to the ML estimator of [11].

The exact evaluation of Eq. (12) requires a high algorithmic complexity since there exists no simple closed-form solution. Therefore, we use an *iterative* procedure to obtain \hat{a} : In an initial step ($j = 0$), a guess for the amplitude $\hat{A}_r^{(0)}$ is made. In iteration step j , Eq. (12) is solved for $\hat{a}^{(j)}$ with a fixed value $\hat{A}_r^{(j-1)}$. Afterwards, Eq. (12) is solved with the obtained value $\hat{a}^{(j)}$ to gain the new estimate $\hat{A}_r^{(j)}$ and so on until no further improvements are achieved. This iterative scheme is suboptimal in comparison to an exact solution, but it provides good results in practice as shown later.

If the interfering noise is not too strong, the value for A_r can also be estimated by taking the mean

$$\hat{A}_r = \sqrt{\frac{1}{L} \sum_{i=0}^L y^2(i)}. \quad (13)$$

The value for L should cover a period of about 20 ms or less so that the sound decay has no significant influence. By this, the RT can be calculated directly by Eq. (12), termed as *non-iterative* GML RTE.

An important special case is given, if no additive noise is (assumed to be) present, i.e., $\sigma_n = 0$. In this case, it can be shown that the LLF of Eq. (11) simplifies to

$$\mathcal{L} = -\frac{N}{2} \left((N-1) \ln(a) + \ln \left(\frac{2\pi}{N} \sum_{i=0}^{N-1} a^{-2i} y^2(i) \right) + 1 \right) \quad (14)$$

so that the parameter \hat{A}_r drops out and only the parameter \hat{a} needs to be determined. In this case, the generalized ML estimator simplifies to the ML estimator of [11].

The devised iterative and non-iterative GML estimator have been used to determine the RT out of a measured RIR disturbed by additive Gaussian noise shown in Fig. 1 ($f_s = T_s^{-1} = 16$ kHz). For comparison, the Schroeder method [6] and the ML approach of Ratnam et al. [11] have been used (see also [14]). The obtained results are compiled in Table 1. For the original (noiseless) RIR, the results of all ML estimators are identical as the GML

approach	RIR	
	noiseless	noisy
Schroeder method [6]	0.97 s	3.20 s
ML RTE [11]	1.01 s	2.00 s
iterative GML RTE	1.01 s	1.03 s
non-iterative GML RTE	1.01 s	1.07 s

Table 1: RTs determined by different estimation methods from sequences shown in Fig. 1.

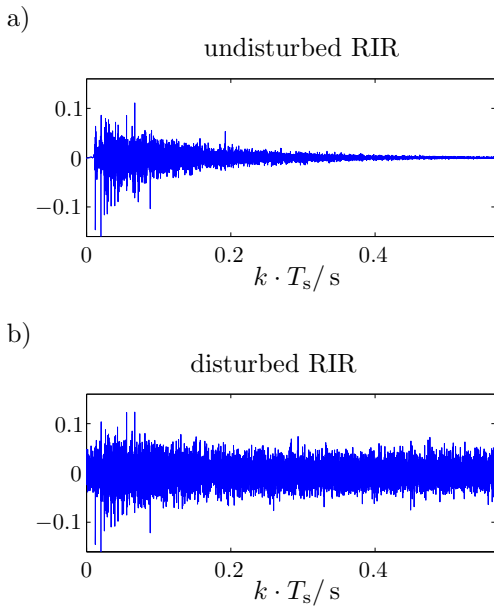


Figure 1: Measured RIR distorted by adding white Gaussian noise ($\sigma_n = 0.02$, $f_s = 16$ kHz).

approach reduces to the simple ML RTE according to Eq. (14). It can also be observed that the result of the ML estimation is very close to the RT obtained by the Schroeder method. For the noisy RIR, only the two GML RTEs achieve satisfactory results despite the strong noise, while the Schroeder method and the simple ML RTE fail completely in this case.

Blind RT Estimation

The ML estimation can also be used for a blind RTE from noisy and reverberant speech signals. It turned out that the direct estimation of the RT from a noisy and reverberant signal is difficult to perform. Instead, it is feasible to denoise the degraded speech signal first. This can be achieved by common speech enhancement techniques such as spectral subtraction or Wiener filtering, cf., [17]. It is important to notice that such methods achieve only a *partial* noise reduction so that residual noise still remains.

Afterwards, a blind RTE is performed by ML estimation and order-statistics filtering similar to the approach of [11]. The ML estimation of Eq. (14) is performed at intervals of R sample instances to a frame $y(\lambda R - N + 1 + i)$ with $\lambda = \lfloor k/R \rfloor$ and $i = 0, 1, \dots, N - 1$. A correct RT estimate can be obtained, if the current segment captures a free decay period following the (sharp) offset of a speech sound. Otherwise, an incorrect RT is obtained, e.g., for segments with ongoing speech, speech onsets or gradually declining speech offsets. Such estimates can be expected to overestimate the RT since the damping of sound cannot occur at a rate faster than the free decay. However, taking the minimum of the last K_1 ML estimates is likely to underestimate the RT since the estimation procedure is a stochastic process. A more robust strategy is to build the histogram of the last K_1 ML estimates and to take the first local maximum $\hat{T}_{60}^{(\text{peak})}(\lambda)$ as final RT estimate, an approach known

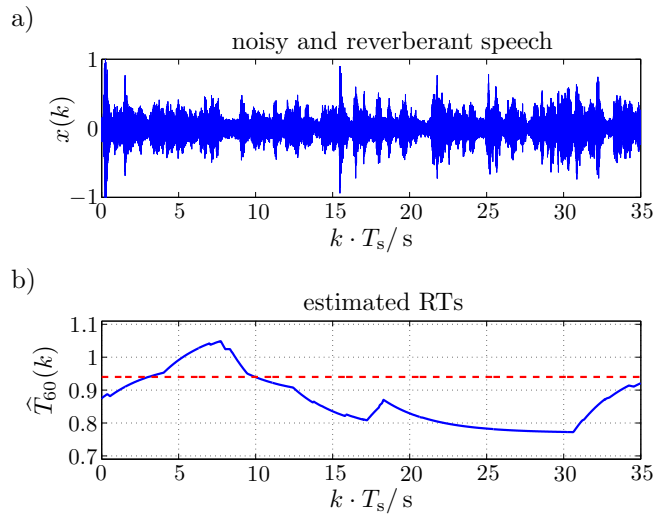


Figure 2: Adaptive, blind estimation of the RT from a reverberant speech signal ($T_{60} = 0.94$ s) distorted by babble noise (SNR = 5 dB, $f_s = 16$ kHz).

as order-statistics filtering. The effects of outliers are efficiently reduced by recursive smoothing

$$\hat{T}_{60}(\lambda) = \beta \hat{T}_{60}(\lambda - 1) + (1 - \beta) \hat{T}_{60}^{(\text{peak})}(\lambda) \quad (15)$$

with $0.9 < \beta < 1$. In contrast to the approach of [13], this blind estimation makes no assumption about the statistical properties of the reverberant (subband) signals (e.g., negative-side variance) and needs thus no calibration procedure. Instead, the presented algorithm exploits the fact that the observed signal contains occasionally small pauses of some hundred milliseconds, which is always fulfilled for speech signals. In contrast to the algorithm of [12], it is also possible to estimate larger RTs ($T_{60} > 0.6$ s).

The devised blind RT estimation has been applied to a noisy speech signal $x(k)$ as shown in Fig. 2. The speech signal is first convolved with the RIR plotted in Fig. 1-a and then distorted by adding babble noise taken from the NOISEX-92 database (see Fig. 2-a). The denoising has been performed by the spectral subtraction rule based on a noise power estimation by means of minimum statistics, cf., [17]. The histogram for the blind RTE is determined by the 400 most recent ML estimates for a bin size (resolution) of 0.11 s. For the ML estimation, a time span of 0.19 s and a frame shift of 0.025 s are taken. A factor of $\beta = 0.995$ is used for Eq. (15).

Fig. 2-b shows that the devised blind RTE achieves an error of about ± 0.16 s. Such an estimation accuracy is usually sufficient for speech enhancement applications outlined in the following.

Application to Speech Enhancement

Spectral enhancement of reverberant *and* noisy speech relies on the following time-domain model: The distorted signal $x(k)$ is given by a superposition of reverberant

speech $z(k)$ and additive noise $v(k)$:

$$\begin{aligned} x(k) &= z(k) + v(k) \\ &= \sum_{n=0}^{L_R-1} s(k-n) h_R(n, k) + v(k) \end{aligned} \quad (16)$$

with $h_R(n, k)$ denoting the time-varying RIR. The reverberant speech can be decomposed into its early and late reverberant speech components according to

$$z(k) = \underbrace{\sum_{n=0}^{L_e-1} s(k-n) h_R(n, k)}_{\text{early rev. speech } z_e(k)} + \underbrace{\sum_{n=L_e}^{L_R-1} s(k-n) h_R(n, k)}_{\text{late rev. speech } z_l(k)}. \quad (17)$$

The suppression of *late reverberant speech* $z_l(k)$ and additive noise $v(k)$ is then accomplished by modeling them both as uncorrelated, random noise processes so that (common) spectral enhancement techniques can be applied [3, 4]. The estimation of the needed spectral variances of the late reverberant speech requires (only) an estimate of the (frequency dependent) RT out of a noisy and reverberant speech signal. In [4], it is assumed that the needed RT can be reliably estimated. In [3], the RT is estimated by a heuristic speech offset detection which, however, is less suitable for noisy speech. In contrast, the blind RTE described before is also suitable for noisy, reverberant speech. This can be used for speech enhancement and allows to achieve a subjective speech quality which is almost equal to the quality achieved by using the actual RT as shown in [15].

Conclusions

A method to estimate the reverberation time (RT) by means of a generalized maximum likelihood (GML) approach is presented. It is derived from a statistical model for the sound decay in reverberant rooms which considers explicitly distortions due to additive noise. The new approach allows to estimate the RT from a measured room impulse response or sound decay distorted by background or measurement noise. The needed noise power estimate can be easily obtained from the observed sequence. The other model parameters (damping factor a and amplitude A_r) can be calculated by an iterative or non-iterative procedure.

The ML estimation can also be used for a blind estimation of the RT from a reverberant and noisy speech signal. After applying a conventional noise reduction system, the RT is estimated by means of a continuous ML estimation followed by order-statistics filtering to select the most likely RT estimate. This new blind RT estimator can achieve an accuracy of less than ± 0.2 s, which is sufficient for the enhancement of noisy and reverberant speech by means of spectral enhancement methods.

Acknowledgment - Thanks to the Institute of Technical Acoustics of RWTH Aachen University for providing the measured RIRs.

References

- [1] H. Kuttruff, *Room Acoustics*, Taylor & Francis, London, 4th edition, 2000.
- [2] J. B. Allen, "Effects of Small Room Reverberation on Subjective Preference," *Journal of the Acoustical Society of America*, vol. 71, no. S1, pp. S5, 1982.
- [3] 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.
- [4] E. A. P. Habets, *Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement*, Ph.D. thesis, Eindhoven University, Eindhoven, The Netherlands, June 2007.
- [5] ISO-3382, "Acoustics-Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters," International Organization for Standardization, Geneva, Switzerland, 1997.
- [6] M. R. Schroeder, "New Method of Measuring Reverberation Time," *Journal of the Acoustical Society of America*, vol. 37, pp. 409–412, 1965.
- [7] W. T. Chu, "Comparison of Reverberation Measurements Using Schroeder's Impulse Method and Decay-Curve Averaging Method," *Journal of the Acoustical Society of America*, vol. 63, no. 5, pp. 1444–1450, May 1978.
- [8] N. Xiang, "Evaluation of Reverberation Times Using a Nonlinear Regression Approach," *Journal of the Acoustical Society of America*, vol. 98, no. 4, pp. 2112–2121, Oct. 1995.
- [9] 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.
- [10] S. Vesa and A. Härmä, "Automatic Estimation of Reverberation Time From Binaural Signals," in *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia (Pennsylvania), USA, Mar. 2005, vol. 3, pp. 281–284.
- [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] 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.
- [13] 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.
- [14] 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.
- [15] H. W. Löllmann and P. Vary, "A Blind Speech Enhancement Algorithm for the Suppression of Late Reverberation and Noise," in *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 2009.
- [16] A. Papoulis and S. U. Pillai, *Probabilities, Random Variables and Stochastic Processes*, McGraw-Hill, New York, 4th edition, 2002.
- [17] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*, Wiley, Chichester, 2006.