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

Postprint

This is the accepted version of a paper presented at *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*.

Citation for the original published paper:

Schüldt, C., Händel, P. (2013)

Blind low-complexity estimation of reverberation time.

In: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, October 20-23, 2013, New Paltz, NY.*

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-136728>

BLIND LOW-COMPLEXITY ESTIMATION OF REVERBERATION TIME

Christian Schüldt and Peter Händel

Royal Institute of Technology
Signal Processing Lab
SE-100 44 Stockholm, Sweden,
crilla@kth.se, ph@kth.se

ABSTRACT

Real-time blind reverberation time estimation is of interest in speech enhancement techniques such as e.g. dereverberation and microphone beamforming. Advances in this field have been made where the diffusive reverberation tail is modeled and the decay rate is estimated using a maximum-likelihood approach. Various methods for reducing the computational complexity have also been presented.

This paper proposes a method for even further computational complexity reduction, by more than 60% in some cases, and it is shown through simulations that the results of the proposed method are very similar to that of the original.

Index Terms— Reverberation time estimation, blind estimation, maximum-likelihood, low-complexity

1. INTRODUCTION

Knowledge of the *reverberation time* T_{60} of an enclosed space, defined as the time required for the reflected (reverberant) sound energy to decay 60 dB below that of the direct sound [1], can be important in many real-time applications, such as for spectral-subtraction-based dereverberation/speech enhancement [2, 3] or dynamically selecting which type of microphone beamforming to use [4]. The parameters determining the reverberation time are the reflectiveness of the surfaces, size and shape of the enclosure, together with possible reflective and/or damping objects (e.g. furnitures in a living room).

Estimating the reverberation time can easily be done by measuring the decay curve slope of a recorded signal directly after an interrupted excitation signal in the form of broad- or narrow-band noise. Because of noise fluctuations, the decay curve will differ from trial to trial, requiring averaging over multiple trials to obtain a reliable estimate. However, it has been shown that the room impulse response (RIR) is related to the ensemble average of different interrupted noise trials and a method, commonly denoted the *Schroeder method* [5], obtains the reverberation time directly from a measured RIR, avoiding multiple trials.

Unfortunately, in many situations it is not possible to use known excitation noise to obtain RIR-measurements and the only information available is a microphone signal containing reverberant speech. Methods that solely use a reverberant speech signal for estimating reverberation time are denoted *blind* reverberation time estimators. Several blind T_{60} estimators have been proposed. In [6], a frequency domain approach is presented, requiring prior training for parameter calibration. In [7], modulation spectral insights are used to develop an adaptive measure termed speech to reverberation modulation energy ratio. It is shown that the inverse of this

ratio is highly correlated with the reverberation time. An alternative approach based on *maximum-likelihood estimation* (ML estimation) in the time domain is presented in [8], where the diffusive tail of reverberation is modeled as an exponentially damped Gaussian white noise process. Approaches for reducing the computational complexity of this method have also been proposed [9, 10].

In [11], performance comparisons between the T_{60} estimators in [6, 7, 10] are made. It was concluded that in general, all three methods are able to estimate the reverberation time to within ± 0.2 s for $T_{60} < 0.8$ s and signal-to-noise ratio ≥ 30 dB. It was also shown that the method in [10] requires significantly lower computational resources compared to the other two. This paper presents a method for even further reduction of the computational complexity.

2. MAXIMUM-LIKELIHOOD ESTIMATION OF THE DECAY RATE

A (sampled) reverberant speech signal $s(k)$ is considered, where sections of beginning speech pauses are modeled as $s(k) = v(k)e^{-\rho k}$, where $k \geq 0$ is the sample index, ρ is the decay rate (corresponding to the reverberation time) and $v(k)$ is a random variable [8, 10]. Using the expression for the instantaneous power

$$E\{s^2(k)\} = E\{v^2(k)\}e^{-2\rho k}, \quad (1)$$

where $E\{\cdot\}$ denotes expectation, the assumption of a diffuse sound field and a source-microphone distance greater than the critical distance [12], the relation between ρ and the reverberation time T_{60} can be expressed as [8]

$$T_{60} = \frac{3}{\rho F_s \log_{10}(e)} \approx \frac{6.91}{\rho F_s}, \quad (2)$$

where F_s is the sampling frequency.

In most publications, e.g. [8, 9, 10, 2], the sequence $v(k)$ is assumed to be i.i.d. Gaussian. In [13], on the other hand, a Laplace distribution was assumed, motivated by the observation that the distribution pattern for temporal decay of a reverberant hand clap is closer to the Laplace distribution. It was shown that the simulated results are very similar, regardless of the Laplace/Gaussian distribution assumption.

In this paper, the Laplacian assumption is used as this will provide better conditions for polynomial approximation, which will be shown in the end of this section. However, if $v(k) \sim \text{Laplace}(0, 1/\lambda)$, then $|v(k)| \sim \text{Exponential}(\lambda)$. So, according to the above model, $|s(k)|$ is represented by a random variable with exponential probability density function (PDF)

$$p_{|s(k)|}(x) = \begin{cases} \lambda e^{\rho k} e^{-\lambda e^{\rho k} x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

An observed sequence $|s(k)|$, with $k \in \{0, \dots, N-1\}$, is modeled by N uncorrelated random variables with *non-identical* PDFs. The joint PDF (likelihood function) of the observed sequence is then

$$\mathcal{L}(\rho, \lambda) = \lambda^N e^{\rho(N-1)N/2} e^{-\lambda \sum_{n=0}^{N-1} e^{\rho n} |s(n)|}, \quad (4)$$

yielding the log-likelihood function

$$\ln \mathcal{L}(\rho, \lambda) = N \ln(\lambda) + \rho(N-1)N/2 - \lambda \sum_{n=0}^{N-1} e^{\rho n} |s(n)|. \quad (5)$$

To find the maximum of (5), the expression is first derived with respect to λ and ρ , resulting in

$$\frac{\partial \ln \mathcal{L}(\rho, \lambda)}{\partial \rho} = N(N-1)/2 - \lambda \sum_{n=0}^{N-1} n e^{\rho n} |s(n)|, \quad (6a)$$

$$\frac{\partial \ln \mathcal{L}(\rho, \lambda)}{\partial \lambda} = N/\lambda - \sum_{n=0}^{N-1} e^{\rho n} |s(n)|. \quad (6b)$$

The derivatives in (6a) and (6b) are then each set to 0 and combined, giving the expression

$$\frac{\sum_{n=0}^{N-1} n e^{\rho n} |s(n)|}{\sum_{n=0}^{N-1} e^{\rho n} |s(n)|} = \frac{N-1}{2}. \quad (7)$$

(It should be noted that (7) also can be obtained by some basic algebraic manipulation of the corresponding expression for the Laplace distributed assumption in [13].)

The ML estimation of ρ is obtained by solving (7). However, as for the corresponding expression for the Gaussian assumption [8, 10], equation (7) is a transcendental equation that has to be solved numerically. In [8], a combined approach using first a bisection method and then Newton-Raphson was used to solve for ρ , requiring around 10–12 iterations in total to obtain a root with reasonable precision. Two faster methods were presented in [9]; one based on a recursive technique for calculating consecutive solutions for ρ and the other a 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 [10, 13], a quantization approach was used where the log-likelihood function was simply evaluated for a fixed number of values of ρ , and the ρ yielding the maximum log-likelihood was selected as the root.

2.1. Solving for the decay rate

In this paper, polynomial approximation is used to obtain a fast and accurate estimation of the decay rate ρ . From (6b) set to 0, the rate of the assumed exponential distribution is given by

$$\lambda = \frac{N}{\sum_{n=0}^{N-1} e^{\rho n} |s(n)|}. \quad (8)$$

Inserting (8) into (5) yields

$$\ln \mathcal{L}(\rho) = N (\ln(N) - f(\rho) + \rho(N-1)/2 - 1), \quad (9)$$

where

$$f(\rho) = \ln \left(\sum_{n=0}^{N-1} e^{\rho n} |s(n)| \right). \quad (10)$$

Figure 1: a) Log-likelihood for different sections of beginning speech pauses for different reverberation times, with the assumption of exponential distribution (see text). The dashed lines are 2nd degree polynomial approximations. b) Difference between the true log-likelihood function and the approximation.

First of all, it can be noted that (10) is monotonically increasing as ρ increases. Secondly, the interesting region of ρ , assuming the sampling frequency $F_s = 8$ kHz, is around $0 < \rho \leq 0.01$, where in this case $\rho = 0.01$ corresponds to a reverberation time of 0.086 seconds and as $\rho \rightarrow 0^+$ the reverberation time goes to infinity (see (2)). Within this region, $f(\rho)$ will exhibit a fairly low degree of curvature and can be approximated by a 2nd degree polynomial. This means that also (9) can be approximated by a 2nd degree polynomial, i.e. $\ln \mathcal{L}(\rho) = c_2 \rho^2 + c_1 \rho + c_0$, and the maximum (here denoted the ML Polynomial (MLP) estimate) is located at $\rho_{MLP} = -c_1/(2c_2)$.

In figure 1 a), equation (9) is plotted for five different speech sections subjected to different amount of reverberation. Also shown in the figure are 2nd degree polynomial approximations (dashed lines). Figure 1 b) shows the difference (error) between (9) and the polynomial approximation. As can be seen in the figure, the polynomials approximates (9) reasonably well, giving an error within ± 20 for $0 \leq \rho \leq 0.09$.

As a comparison, the corresponding expression to (9), assuming a Gaussian distribution of $v(k)$, is [8, 10]

$$\ln \mathcal{L}_G(\rho) = \frac{N}{2} (\ln(N) - \ln(2\pi) + \rho(N-1) - g(\rho) - 1), \quad (11)$$

where

$$g(\rho) = \ln \left(\sum_{n=0}^{N-1} e^{2\rho n} s^2(n) \right). \quad (12)$$

The difference between (12) and (10) is the squaring of the terms, which slightly increases the steepness of the resulting log-likelihood function (11), compared to (9). Thus, approximating (9) with a 2nd degree polynomial is likely to give lower error than approximating (11) with a 2nd degree polynomial. In figure 2 a), equation (11) is plotted for five different speech sections subjected to

Figure 2: a) Log-likelihood for different sections of beginning speech pauses for different reverberation times, with the assumption of Gaussian distribution (see text). The dashed lines are 2nd degree polynomial approximations. b) Difference between the true log-likelihood function and the approximation.

different amount of reverberation together with the corresponding polynomial approximations (dashed lines). Figure 2 b) shows the difference (error) between (11) and the polynomial approximation. It can clearly be seen that in this case the 2nd degree polynomial approximation of the likelihood function using the exponential distribution assumption yields significantly lower approximation error compared to the 2nd degree polynomial approximation of the likelihood function using the Gaussian assumption.

2.1.1. A note on the computational complexity

As N is typically in the order of 1000 [9], it is clear that the most demanding part of the described approach, in terms of computational complexity, is the evaluation of the sum in (10), requiring $2N$ multiplications, N additions, N magnitude calculations, one evaluation of e^ρ and one natural logarithm calculation. The polynomial approximation described in the previous section requires a minimum of three evaluations of (10). As a comparison, the fast block based approach in [9] requires $N(3R + 1) + 2R$ multiplications and $(2N - 1)R$ additions, where R is the number of iterations. For e.g. $R = 5$ it is obvious that the proposed polynomial approximation can give significantly lower computational complexity; in some cases more than 60% lower depending on N . It is also obvious that the proposed approach is significantly faster than the direct quantization method used in [10, 13] and it should be noted that in contrast to this method, the proposed method is not in the same way restricted to a limited pre-defined set of estimates.

3. FINDING SECTIONS OF BEGINNING SPEECH PAUSES

The assumptions used for T_{60} estimation described in section 2 are only valid if the considered frame $|s(k)|$ where $k \in \{0, \dots, N-1\}$

D	M	M_Δ	$l_{\min}$	K_s	α
2	1911	137	3	400	0.996

Table 1: Parameters used by the T_{60} estimation methods.

corresponds to the beginning of a speech pause. To detect such frames, the approach presented in [10], except the second histogram step for fast tracking of rapid T_{60} -changes, is used in this paper. The method is described briefly below (the reader is referred to [10] for a more detailed description).

The reverberant speech signal is downsampled by a factor D and segmented into frames of length M with M_Δ samples overlap. Each frame is then divided into L sub-frames. Then it is checked, for all sub-frames, whether the energy, maximum and minimum value deviates from the successive sub-frame according to

$$\begin{aligned} \psi_E(t, l) &> \psi_E(t, l+1), \\ \psi_{\max}(t, l) &> \psi_{\max}(t, l+1), \\ \psi_{\min}(t, l) &< \psi_{\min}(t, l+1), \end{aligned} \quad (13)$$

where $\psi_E(t, l)$ is the sub-frame energy, $\psi_{\max}(t, l)$ is the maximum sub-frame sample, $\psi_{\min}(t, l)$ is the minimum sub-frame sample, l is the sub-frame index and t is the frame index. If any of the conditions in (13) are false, it is checked if $l \geq l_{\min}$ and if so, the comparison is aborted and the next frame is processed. Otherwise, the consecutive sub-frames for which the conditions in (13) are true are combined and detected as a possible sound decay. For this segment, the reverberation time is estimated according to the procedure described in section 2.

The reverberation time of the last K_s estimates are kept in a histogram and the location of the maximum of the histogram taken as the “current” T_{60} estimate, denoted $\hat{T}_{60}(t)$. To reduce the variance, the final estimate $\tilde{T}_{60}(t)$ is calculated by recursive smoothing according to

$$\hat{T}_{60}(t) = \alpha \hat{T}_{60}(t-1) + (1-\alpha) \tilde{T}_{60}(t), \quad (14)$$

where α is a forgetting factor.

4. SIMULATIONS

The performance of the proposed algorithm (MLP) was compared to the one in [10] (here denoted the ML approach with Gaussian assumption (MLG)) as well as the one in [13] (here denoted the ML approach with Laplacian assumption (MLL)), and the Schroeder method [5] was used as reference. Schroeder’s method used least squares fitting between the decay ranges -5 to -35 dB [8]. The histogram (see section 3) used by the ML methods was set to contain 11 bins ranging from 0.2 s to 1.2 s. Other parameters for the different methods were also set to the same values, see table 1. For the proposed method, the polynomial approximations were calculated by evaluating (9) in three points, $\ln \mathcal{L}(i)$ where $i \in \{-0.0001, 0.001, 0.0025\}$, and then solving a 3×3 linear system to obtain the polynomial coefficients c_0, c_1 & c_2 . The ML estimate was obtained as $\rho_{\text{MLP}} = -c_1/(2c_2)$ (see section 2.1).

A speech file, with sampling frequency 16 kHz, containing 24 utterances from different speakers (both male and female) recorded in an anechoic room was convoluted with 20 different RIRs measured in rooms with different reverberation time. The RIRs were taken from the AIR database [14] and downsampled to 16 kHz. The reverberant speech was used as input and the mean T_{60} taken over all considered frames was compared for all three methods, respectively.

Figure 3: a) Comparison of MLG [10] and MLL [13] with the proposed low-complexity approach MLP. Schroeder's method is used as ground-truth. The dashed diagonal line is shown for reference. b) Error variance. The reverberation time is increasing with the RIR index.

5. RESULTS

The results are shown in figures 3 a) and 3 b), where the mean reverberation time for the compared ML methods are plotted against the reverberation time obtained by Schroeder's method for the 20 different RIRs and the respective error variances are shown, respectively. It can be seen that the assumption of Laplace distributed samples (i.e. exponentially distributed sample magnitudes) (MLL) and the assumption of Gaussian distributed samples (MLG) yield very similar results, which is in agreement with what is reported in [13]. It can also be seen that the T_{60} estimates of the proposed low-complexity method are close to those of the original methods and that all estimates correspond fairly well to those obtained by Schroeder's method. All methods also show similar estimation variance for the simulation set, appearing to increase for reverberation times close to 1 s.

6. CONCLUSIONS

A low-complexity method for blind estimation of reverberation time has been presented. The method builds upon that of [10], assuming, instead of Gaussian distribution, an exponential distribution of the magnitude of the samples in the the diffusive reverberation and it has been shown that this allows the efficient 2nd degree polynomial approximation for obtaining an estimate of the ML T_{60} -solution, yielding a significantly lower computational complexity compared to previous methods. It was shown through simulations with RIRs obtained from measurements that the T_{60} estimates of the proposed method correspond well with those obtained by the methods in [10] and [13].

Further work will be more extensive evaluation (theoretical as well as with additional simulations), and comparison of the results to a straight-forward approach of simply taking the natural logarithm of each squared sample and estimating ρ through linear regression [3]. Another possible extension of the work would be to estimate the reverberation time in complex sub-bands, using the

property that the squared magnitude of a complex Gaussian distributed random variable, assuming that the real and imaginary part are uncorrelated, will have exponential distribution.

7. REFERENCES

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] M. R. Schroeder, "New method of measuring reverberation time," *Journal of the Acoustical Society of America*, vol. 37, no. 3, pp. 409–412, 1965.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] H. Kuttruff, *Room Acoustics*, 4th ed. Taylor & Francis, Oct. 2000.
- [13] J. Tariquallah 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.
- [14] M. Jeub, M. Schafer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *Proc. of the 16th Intl. Conf. on Digital Sig. Proc.*, July 2009.