Fast Algorithms for Blind Estimation of Reverberation Time

Rama Ratnam, Douglas L. Jones, *Fellow, IEEE*, and William D. O'Brien, Jr., *Fellow, IEEE*

***Abstract*—The reverberation time (RT) measures the persistence of a sound in enclosed acoustic spaces. In a previous work, a method for the blind maximum-likelihood estimation (MLE) of RT using passively received microphone signals was presented. The procedure overcomes the drawbacks of current methods that use a controlled sound source for RT determination. Here, fast algorithms for online implementation of the method are developed. One algorithm, suitable for a one-time determination of the RT, requires $O(N)$ computations for a data frame of length N. A second IIR algorithm, based on Q-levels of quantization, requires $O(Q)$ computations. Results for speech data and choice of algorithms are discussed.**

***Index Terms*—Acoustic reverberation time, blind estimation, fast algorithms.**

I. Introduction

Room reverberation affects the perception of important acoustic signals, such as speech, by introducing spectral coloration. In the process, the waveform carrier (fine structure) and its envelope become distorted. Speech intelligibility and voice quality are determined by the fine structure and envelope of speech [1]; thus, when phonemes are blurred or masked, intelligibility is affected [2]. For many important applications, such as hearing aids and hands-free telephony, an estimate of room reverberation time can be used to choose a signal-processing strategy most appropriate to the environment. For example, one might switch from fixed beamforming in highly reverberant environments to adaptive beamforming in low-to-moderate reverberation [3]–[5]. However, in these applications, the environment can change dynamically due to sensor movement or a change in room geometry. Thus, it would be of practical importance to quantify room reverberation in real-time, using passively received microphone signals.

The parameter that is most often used to characterize reverberation is the reverberation time (RT). This is defined as the time taken by a sound to decay 60 dB below the initial level, after it has been switched off. It is equivalent to assuming an exponentially decaying envelope with time-constant τ, and so, RT is linearly related to the time-constant $(\mathrm{RT} = 6.91\tau)$. RT can be determined either analytically, using the known geometry and absorptive characteristics of the acoustic space [6], or by radiating a test sound into the enclosure [7]. In the latter, an impulsive sound or a burst of noise is generated, and the free-decay is tracked after the sound ceases. When the room geometry is unknown or the sound source is unknown or cannot be controlled, the methods cannot be applied. Earlier we developed a blind estimation procedure for tracking the room time-constant from passively received sounds [8]. The system consists of a maximum-likelihood estimation procedure that tracks the decay time-constant of the sounds. Because the received sounds are unknown and may be composed of segments (as with speech), only a subset of estimates are true estimates of the room RT. These are the estimates obtained from regions of free-decay between sound segments. The remaining estimates are spurious, reflecting the fluctuations in the ongoing sound envelope. The estimates corresponding to the free-decay are then extracted from the accumulated estimates using an order-statistics filter. While the procedure has been validated with experimental data from a number of listening environments [8], it did not however, suggest how the method could be implemented efficiently in online situations. The work reported here presents two novel algorithms for fast online implementation in digital signal processors. Accuracy and speed requirements are discussed, and criteria for choice of algorithms are proposed.

Manuscript received May 23, 2003; revised September 12, 2003. The work was supported by grants from the National Institutes of Health (R21DC04840), Phonak AG, and The Beckman Institute. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Erchin Serpedin.

R. Ratnam is with the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: ratnam@uiuc.edu).

D. L. Jones and W. D. O'Brien, Jr., are with the Department of Electrical and Computer Engineering and the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA.

Digital Object Identifier 10.1109/LSP.2004.826667

II. Maximum-Likelihood Estimation of Reverberation Time

When a sound that is radiated into an enclosure is abruptly switched off, it persists for some time. First there is the direct sound (due to delay induced by the transmission path), followed by a series of early reflections, and then a reverberant tail that consists of dense reflections due to multiple scattering. The dense reflections are referred to as reverberation and they are likely to persist from tenths of seconds to several seconds. Unlike the direct sound and early reflections, acoustic reverberation consists of a fine structure (the carrier waveform of the decaying envelope) that can be described only statistically [6], [7]. Usually, the fine structure is considered to be an uncorrelated random process (see [7] for a discussion), however, the decaying envelope is a deterministic signal parametrized by a time-constant τ. This decay time-constant is linearly proportional to the reverberation time [6].

Let the fine structure of the reverberant tail be denoted by a random sequence $x(n)$, $n \geq 0$, of independent and identically distributed random variables drawn from the normal distribution $\mathcal{N}(0, \sigma)$. Further, we define a deterministic constant

$a = \exp(-1/\tau)$, where τ is the time-constant to be estimated. We note that $a \in [0, 1)$ maps one-to-one onto $\tau \in [0, \infty)$. The model for room decay is a product of the fine structure and the deterministic envelope, and so the observations y are specified by a geometrically decaying sequence $y(n) = a^n x(n)$. Due to the time-varying term a^n, the $y(n)$ are independent but not identically distributed, and their probability density function is $\mathcal{N}(0, \sigma a^n)$. For purposes of estimating the room decay time, we consider a finite sequence of observations, $n = 0, \ldots, N$ where N will be referred to as the estimation interval or estimation window length. For notational simplicity, denote the N-dimensional vector of observations y by $\mathbf{y}$. Then the likelihood function of $\mathbf{y}$ (the joint probability density), parameterized by a and σ, is

$$P(\mathbf{y}; a, \sigma) = \left(\frac{1}{2\pi a^{(N-1)}\sigma^2}\right)^{\frac{N}{2}} \exp\left(-\frac{\sum_{n=0}^{N-1} a^{-2n} y^2(n)}{2\sigma^2}\right). \tag{1}$$

For a fixed observation window N and a sequence of observations $y(n)$, the likelihood function is parameterized solely by the time-constant a and the diffusive power σ.

Given the likelihood function, the parameters a and σ can be estimated using a maximum-likelihood approach. First, we take the logarithm of (1) to obtain the log-likelihood function

$$L(\mathbf{y}; a, \sigma) = -\frac{N(N-1)}{2}\ln(a) - \frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1} a^{-2n} y^2(n). \tag{2}$$

The partial derivatives of the log-likelihood function are

$$\frac{\partial \ln L(\mathbf{y}; a, \sigma)}{\partial a} = -\frac{N(N-1)}{2a} + \frac{1}{a\sigma^2}\sum_{n=0}^{N-1} n a^{-2n} y^2(n), \tag{3}$$

$$\frac{\partial \ln L(\mathbf{y}; a, \sigma)}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_{n=0}^{N-1} a^{-2n} y^2(n). \tag{4}$$

The maximum-likelihood estimates of a and σ are obtained in the usual manner by setting the partial derivatives of the log-likelihood function, (3) and (4), to zero. This results in the following pair of equations for a and σ, respectively

$$0 = -\frac{N(N-1)}{2a} + \frac{1}{a\sigma^2}\sum_{n=0}^{N-1} n a^{-2n} y^2(n), \tag{5}$$

$$\sigma^2 = \frac{1}{N}\sum_{n=0}^{N-1} a^{-2n} y^2(n). \tag{6}$$

Note that (5) is an implicit expression for a and must be solved iteratively, whereas (6) provides the ML estimate of σ directly if a is known.

III. Determining the Most Likely Room Time Constant

The estimation proceeds by sliding the window of length N over the incoming sequence of observations y, producing a sequence of estimates at intervals determined by the step size of the advancing window. Estimates of a are accumulated in a histogram from which the room time-constant must be inferred. Because the estimator tracks sound decay, the histogram will accumulate the time-constant due to fluctuations in the ongoing sound envelope, in addition to the free decay occurring after a sound offset. The former estimates are spurious and must be rejected, because they reflect the properties of the ongoing sound envelope rather than the reverberation curve. We formulate a decision-making strategy by noting that no segment of sound can decay at a rate faster than a free decay, for the free decay occurs when the sound offset is instantaneous. All other decays are a convolution of the ongoing sound envelope with the room impulse response, and hence their time-constants exceed that of the room. We construct an order-statistics filter that determines the dominant peak at the lower end of the range. The time-constant correponding to the peak is the room time-constant. The procedure was validated using simulations and real room data, and results are in good agreement with the actual RT values [8].

The variance of the estimates is dependent on the estimation window length N. While close form expressions are not available due to the implicit nature of (5), it is possible to compute the Crámer-Rao bounds for a and σ [8]. If a^* and σ^* are the estimates obtained from a solution of (5) and (6), the lower bounds for the variance of the estimates are $\mathbf{E}[(a^* - a)^2] \geq 6a^2/N(N^2-1)$ and $\mathbf{E}[(\sigma^* - \sigma)^2] \geq \sigma^2(2N-1)/N(N+1)$. Thus, the estimation is asymptotically unbiased and efficient. In practice, the value of N is determined by the duration of gaps (or silent periods) between sound segments and cannot be made arbitrarily large. For speech, the gaps are of the order of hundreds of milliseconds. Estimation windows greater than this can create undesirable end effects where the onset of the following segment can creep into the window. Thus, the value of N must be determined empirically or adjusted adaptively and should not exceed the mean duration of gaps.

The most computationally intensive step is the solution of the likelihood equation for a. We note that $a \in [0, 1)$ maps one-to-one onto $\tau \in [0, \infty)$. Thus, the geometric ratio is highly compressive and values of a in real environments are likely to be close to one. For example, if the sampling frequency is 16 kHz and τ ranges from 0.03 s to ∞, a ranges from 0.9979 to 1. For higher sampling frequencies, the range of a is further compressed. Thus, it is advantageous to estimate a rather than τ because the range of values is tightly bounded. Earlier, the Newton–Raphson procedure was used for solving (5) and (6) [8]. The method was found to converge very slowly when the initial value was far from the solution, and it required the calculation of the score function (3) and its gradient. This affected both speed and computational effort and makes the scheme unsuitable for online implementation.

Two new procedures that reduce computational complexity and are suitable for online implementation are presented. The first does not solve (5) and (6) directly, but uses a fast IIR filter to determine the a value that has maximum likelihood in a range of quantized values. This procedure is compared to a second procedure that employs iterative gradient optimization. While it is slower than the IIR algorithm, it solves the ML equations exactly and is significantly faster than the Newton–Raphson scheme. The new algorithms are compared for computational complexity and speed.

IV. Fast Online Algorithm

For most applications, it is not necessary to determine the time-constant to arbitrary precision. Instead, it is possible to sacrifice accuracy of a by eliminating the iterative solution of (5). This approach is taken by quantizing the range of a (these form the bins of the histogram of a), calculating the likelihood value, and assigning the highest likelihood to that bin in the histogram. In this way, the histogram used by the order-statistics filter is explicitly constructed as part of the algorithm.

Let the range of $a \in [0, 1)$ be quantized into Q values, so that we have $a_j, j = 1, \ldots, Q$. Then, for each a_j, the log-likelihood function given by (2) can be rewritten using the solution for σ (6) as

$$L(a_j; \mathbf{y}) = -\frac{N}{2}\left\{(N-1)\ln(a_j) - \ln\left(\frac{2\pi}{N}\sum_{n=0}^{N-1} a_j^{-2n} y^2(n)\right) - 1\right\}. \quad (7)$$

We select the best estimate of a as

$$a = \arg\max\{L(a_j; \mathbf{y})\}. \quad (8)$$

In practice, the computation of $L(a_j; \mathbf{y})$ on arrival of the next sample requires only a recursive update of the second term inside the parenthesis in (7), thereby reducing the implementation to a first-order IIR filter which is linear in $y^2(n)$. Let $\beta = a^{-2}$, then we define

$$g(n) = \beta^{N-1} \sum_{r=n-N+1}^{n} \beta^{r-n} y^2(r) \quad (9)$$

and define the recursive update rule for $g(n)$ as

$$g(n+1) = \beta^{-1}\left(g(n) + \beta^N y^2(n+1) - y^2(n+1-N)\right). \quad (10)$$

Thus, the update requires 4 MULS and 2 ADDS, assuming that β^{-1} and β^N are pre-calculated. The $\ln(a_j)$ and $\ln(2\pi/n)$ can also be pre-calculated. From (7), we note that for each a_j, we require 3 further ADDS, one MUL and the calculation of $\ln(g(n))$. Thus, for Q bins, we require $5Q$ MULS, $5Q$ ADDS, and the evaluation of Q logarithms. The $g(n)$ are directly proportional to σ, whose order of magnitude may not be easy to guess. Thus, the evaluation of the logarithm using a series or asymptotic expansion may not be accurate. Instead, it may be efficient to perform a table-lookup followed by linear interpolation to calculate $\ln g(n)$.

V. Fast Block Algorithm

A block algorithm that performs a single estimate in the most efficient manner is presented here. The method relies on an iterative solution of (5) using a gradient optimization procedure. The solution for σ is known from (6), and so we can rewrite (3) as

$$\frac{\partial \ln L(\mathbf{y}; a, \sigma)}{\partial a} = \frac{N}{a}\left\{-\frac{(N-1)}{2} + \frac{\sum_{n=0}^{N-1} n a^{-2n} y^2(n)}{\sum_{n=0}^{N-1} a^{-2n} y^2(n)}\right\}. \quad (11)$$

TABLE I
COMPUTATIONAL COST FOR EXACT SOLUTION OF MAXIMUM-LIKELIHOOD EQUATION

	Operation	Multiplications	Additions
0	Initialize $A = -(N-1)/2$		
1	$y^2(n)$	N	0
2	$\beta^n = a^{-2n}$	N	0
3	$\sum_{n=0}^{N-1} \beta^n y^2(n)$	$N-1$	$N-1$
4	$\sum_{n=0}^{N-1} n\beta^n y^2(n)$	$N-1$	$N-2$
5	$N \times \beta \times (A + (4)/(3))$	3	1
6	Gradient update $a \leftarrow a - \mu \times (5)$	1	1
7	Iterations ((2) through (6))	R	
	Total	$N(3R+1) + 2R$	$(2N-1)R$

The update rule for solving the equation is the iterative procedure specified by

$$a_{k+1} = a_k - \frac{\mu \partial \ln L(\mathbf{y}; a, \sigma)}{\partial a} \quad (12)$$

where μ is a fixed parameter that provides an independent means of controlling the speed of convergence. It is determined empirically from an examination of the iterative procedure (12). In the Newton–Raphson scheme the step size is $\partial \ln L/\partial a / \partial^2 \ln L/\partial a^2$, and from (12) it can be seen that $\mu \sim O((\partial^2 \ln L/\partial a^2)^{-1})$. This can be used as a guide for selecting the value of μ.

Although a theoretical investigation is beyond the scope of this work, it was found that the Newton–Raphson method required several hundred iterations to converge, whereas with appropriate tuning of μ, the gradient optimization procedure required about five iterations. The computational costs are summarized in Table I. The frame length N is typically of the order of 10^3, and is therefore much greater than the number of iterations R. Thus, the overall number of operations is $O(N)$.

VI. Experimental Results

Speech consisting of 15 isolated word utterances separated by gaps of 200 ms (total duration of 12 s, sampled at 20 kHz) was convolved with a synthetic room impulse response ($\tau = 72$ ms corresponding to RT $= 0.5$ s, and $\sigma = 1$). The convolved signal was processed by the block and fast online algorithms. Histograms of RT estimates are shown in Fig. 1, with both algorithms demonstrating a strong peak at the true RT. The block algorithm is more accurate (reflected in the smaller spread in the histogram) as it solves the ML equations, but it is computationally intensive. On average, the block algorithm was slower by about two orders of magnitude than the online algorithm using a MATLAB (The MathWorks, Inc.) script.

The preceding analysis suggests that the reverberation-time estimation procedure can be implemented with reasonable computational costs. Our interest here is in the online method which has the dual advantage of providing fast estimates and construction of an online histogram in pre-determined bins.

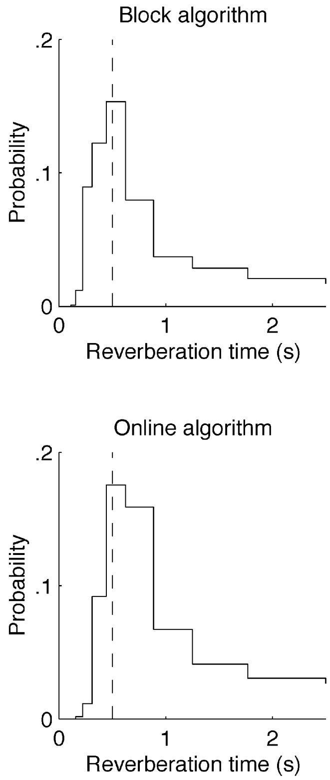

Fig. 1. Histogram of reverberation-time estimates. Top panel shows the block algorithm and bottom panel shows the fast online IIR algorithm. Speech was convolved with a simulated room inpulse response with $\mathrm{RT} = 0.5$ s (dashed line). The histograms peak at the simulated RT value. The block algorithm was run on every sample (i.e., no skips) and the histogram bins were the same as those used in the online algorithm (20 bins, logarithmically spaced in time axis).

The advantage of the online over the block algorithm will be apparent under some, but not all, conditions. Let us say that the block algorithm is run at intervals of M samples. Then from the preceding analysis, the block algorithm performs operations of the order of NR/M ADDS and MULS per sample, whereas the online algorithm will perform order-Q ADDS and MULS. Thus, if $Q \ll NR/M$, then the online method may be preferable to the block algorithm. As an example, let us say that we are sampling at 10 kHz in a moderately reverberant environment ($\tau = 0.1$ s, corresponding to RT of about 0.7 s). The filter length N is of the order of τ, i.e., 1000 samples. If it is necessary to perform about $R = 5$ iterations for the block algorithm, then we will perform about 5000 ADDS and MULS for every block. If we assume that $Q = 10$, then it may be advantageous to use the online algorithm if the block updates are required no more frequently than once every 50 ms. This is an order-of-magniture estimate, and a more detailed analysis of the block algorithm will be required due to the additional overhead of constructing a histogram, choosing the appropriate value of the gradient step size μ, etc.

In practice, the only information that may be needed is whether reverberation is high, moderate or low. Thus, the number of bins Q that are required may not exceed 10. We recommend 2 bins for very high (>10 s) and low values (<0.01 s). Estimates falling within these bins may be rejected, the former suggesting that the algorithm is tracking a region that may not be a decay and the latter suggesting open-air or anechoic conditions. In between these extremes, 5–6 bins at intermediate values can be selected to signify a range of classifications, from low to highly reverberent. Thus, the choice of an algorithm may be dictated by the context, accuracy and frequency of estimates most appropriate for the application.

VII. Conclusion

This work presents fast algorithms for computing the room reverberation time using passively recorded microphone signals. The system is based on a maximum-likelihood estimation procedure followed by an order-statistics filter [8]. In contrast to the standard method [7] that requires control of an active sound source, the current method is a blind procedure that can be applied to passively received microphone signals. An efficient block algorithm requiring $O(N)$ computations was presented for applications requiring a single estimate. This estimation procedure is accurate as it provides independent sampling of the estimated parameters when succesive windows do not overlap. For applications where more frequent estimates are required, such as in microphone arrays or hearing aids, or where the enclosure geometry changes frequently, a fast algorithm requiring significantly fewer computations was also presented. The model assumed in this work provides equal weighting to all estimates. The variance of the estimates can be reduced further if the model were to incorporate an additive regularization term that gives greater weight to the initial portion of the decay. This approach is suggested for future work. Both algorithms are asymptotically equivalent as the estimation window length increases. The algorithms can be implemented as single-channel or multi-channel versions depending on the application. They were tested with real room data using speech sounds and produced estimates in agreement with the actual reverberation times.

Acknowledgment

The authors would like to thank the members of the Intelligent Hearing Aid Project at the Beckman Institute, University of Illinois, for their comments and criticisms. They also thank the anonymous referees whose suggestions have helped to strengthen the work and provided directions for future work.

References

[1] S. Rosen, "Temporal information in speech: acoustic, auditory and linguistic aspects," *Phil. Trans. R. Soc. Lond. B*, vol. 336, pp. 367–373, 1992.

[2] A. K. Nabalek, T. R. Letowski, and F. M. Tucker, "Reverberant overlap- and self-masking in consonant identification," *J. Acoust. Soc. Amer.*, vol. 86, pp. 1259–1265, 1989.

[3] S. Allegro, M. Buechler, and S. Launer, "Automatic sound classification inspired by auditory scene analysis," in *Proc. Eur. Conf. Sig. Proc. (EURASIP)*, 2001.

[4] J. E. Greenberg and P. M. Zurek, "Microphone-array hearing aids," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. Brandstein and D. Ward, Eds. Berlin, Germany: Springer-Verlag, 2001, pp. 229–253.

[5] J. E. Greenberg, J. G. Desloge, and P. M. Zurek, "Evaluation of array-processing algorithms for a headband hearing aid," *J. Acoust. Soc. Amer.*, vol. 113, pp. 1646–1657, 2003.

[6] H. Kuttruff, *Room Acoustics*, 3rd ed. New York: Elsevier, 1991.

[7] M. R. Schroeder, "New method for measuring reverberation time," *J. Acoust. Soc. Amer.*, vol. 37, pp. 409–412, 1965.

[8] R. Ratnam, D. L. Jones, B. C. Wheeler, W. D. O'Brien Jr., C. R. Lansing, and A. S. Feng, "Blind estimation of reverberation time," *J. Acoust. Soc. Amer.*, vol. 114, pp. 2877–2892, 2004.