Blind estimation of reverberation time

Rama Ratnam,[a] Douglas L. Jones, Bruce C. Wheeler, William D. O'Brien, Jr., Charissa R. Lansing, and Albert S. Feng

Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

(Received 14 March 2003; revised 2 August 2003; accepted 18 August 2003)

The reverberation time (RT) is an important parameter for characterizing the quality of an auditory space. Sounds in reverberant environments are subject to coloration. This affects speech intelligibility and sound localization. Many state-of-the-art audio signal processing algorithms, for example in hearing-aids and telephony, are expected to have the ability to characterize the listening environment, and turn on an appropriate processing strategy accordingly. Thus, a method for characterization of room RT based on passively received microphone signals represents an important enabling technology. Current RT estimators, such as Schroeder's method, depend on a controlled sound source, and thus cannot produce an online, blind RT estimate. Here, a method for estimating RT without prior knowledge of sound sources or room geometry is presented. The diffusive tail of reverberation was modeled as an exponentially damped Gaussian white noise process. The time-constant of the decay, which provided a measure of the RT, was estimated using a maximum-likelihood procedure. The estimates were obtained continuously, and an order-statistics filter was used to extract the most likely RT from the accumulated estimates. The procedure was illustrated for connected speech. Results obtained for simulated and real room data are in good agreement with the real RT values.
[DOI: 10.1121/1.1616578]

PACS numbers: 43.66.Yw [MK] Pages: 2877–2892

I. INTRODUCTION

The estimation of room reverberation time (RT) has been of interest to engineers and acousticians for nearly a century (Sabine, 1922; Kuttruff, 1991). The RT of a room specifies the duration for which a sound persists after it has been switched off. The persistence of sound is due to the multiple reflections of sound from the various surfaces within the room. Historically, the RT has been referred to as the T_{60} time, which is the time taken for the sound to decay to 60 dB below its value at cessation.

Reverberation results in temporal and spectral smearing of the sound pattern, thus distorting both the envelope and fine structure of the received sound. Consequently, the RT of a room provides a measure of the listening quality of the room. This is of particular importance in speech perception where it has been noted that speech intelligibility reduces as the RT increases, due to masking within and across phonemes (Knudsen, 1929; Bolt and MacDonald, 1949; Nabelek and Pickett, 1974; Nabelek and Robinson, 1982; Nabelek *et al.*, 1989). The effect of reverberation is most noticeable when speech recorded by microphones is played back via headphones. Previously unnoticed distortions in the sound pattern are now clearly discerned even by normal listeners [see Hartmann (1997) for a discussion], highlighting the remarkable echo suppression and dereverberation capabilities of the normal auditory system when the ears receive sounds directly. For hearing-impaired listeners, the reception of reverberant signals via the microphone of a hearing aid exacerbates the problem of listening in challenging environments. While derverberation is an active area of investigation, state-of-the-art hearing aids, or other audio processing instruments, implement signal processing strategies tailored to specific listening environments. These instruments are expected to have the ability to evaluate the characteristics of the environment, and accordingly turn on the most appropriate signal processing strategy. Thus, a method that can characterize the RT of a room from passively received microphone signals represents an important enabling technology.

In the early 20th century, Sabine (1922) provided an empirical formula for the explicit determination of RT based solely on the geometry of the environment (volume and surface area) and the absorptive characteristics of its surfaces. Since then, Sabine's reverberation-time equation has been extensively modified and its accuracy improved [see Kuttruff (1991) for a historical review of the modifications], so that, today, it finds use in a number of commercial software packages for the acoustic design of interiors. Formulas for calculation of RT are used in anechoic chamber measurements, design of concert halls, classrooms, and other acoustic spaces where the quality of the received sound is of greatest importance, and the extent of reverberation must be controlled. However, to determine the RT of existing environments, both the geometry and the absorptive characteristics have to be first determined. When these cannot be determined easily, it is necessary to search for other methods, such as those based purely on the controlled recordings of excitation sounds radiated into the test enclosure.

Methods that employ an excitation signal for measuring RT are based on sound decay curves. In the Interrupted Noise Method (ISO 3382, 1997), a burst of broad- or

[a]Electronic mail: ratnam@uiuc.edu

narrow-band noise is radiated into the test enclosure. When the sound field attains steady state, the noise source is switched off and the decay curve is recorded. RT is estimated from the slope of the decay curve. However, because of fluctuations in the excitation noise signal, the decay curve will differ from trial to trial, and so RTs from a large number of decay curves must be averaged to obtain a reliable estimate. To overcome this drawback Schroeder (1965a, 1966) developed the Integrated Impulse Response Method where the excitation signal is a brief pulse, either broad- or narrow-band. For a brief pulse the enclosure output is simply the impulse response of the enclosure in the specified frequency band. Schroeder showed that the impulse response of the enclosure is related via a certain integral to the ensemble average of the decay curve obtained using the interrupted noise method, and so repeated trials were unnecessary. Both methods, while theoretically and practically important, require careful control of the experiment. Specifically, a suitable excitation signal must be available, and it must have sufficient power to provide at least a 35-dB decay range before the noise floor is encountered [see ISO 3382 (1997) for specifications of the experiment]. Under these conditions, both methods provide reliable RT estimates, with Schroeder's method being superior because it is the average of an infinite number of interrupted noise measurements.

While Schroeder's method continues to have immense practical utility, and has been improved over the years (see Chu, 1978; Xiang, 1995, for example), there is at present no ''blind'' method that can estimate room RT from passively received microphone signals. The objective of this work is to establish a method for determining RT when the room geometry and absorptive characteristics are unknown, or when a controlled test sound cannot be employed. A blind method that works with speech sounds would be particularly important for incorporating in hearing-aids or hands-free telephony devices. Partially blind methods have been developed in which the characteristics of the room are ''learned'' using neural network approaches (Tahara and Miyajima, 1998; Nannariello and Fricke, 1999; Cox *et al.*, 2001), or some form of segmentation procedure is used for detecting gaps in sounds to allow the sound decay curve to be tracked (Lebart *et al.*, 2001). The only other method that can be described as truly blind is ''blind dereverberation,'' where the aim is to recover a sound source by deconvolving the room output with the unknown room impulse response. When deconvolution is successful, a useful by-product is the room impulse response from which the RT can be estimated (using, say, Schroeder's method). However, deconvolution is difficult to perform because it requires the room impulse response to be minimum phase, a condition that is not met in most real environments (Neely and Allen, 1979; Miyoshi and Kaneda, 1988). It should be noted that RT can always be determined if the room impulse response is known, whether it is minimum or nonminimum phase. The minimum phase condition is only necessary for determining the impulse response via a deconvolution. This limits the applicability of the method.

Here we develop a technique for blind estimation of reverberation time based solely on passively recorded sounds. The estimator is based on a simplified noise decay curve model describing the reverberation characteristics of the enclosure. Sounds in the test enclosure (speech, music, or other pre-existing sounds) are continuously processed and a running estimate of the reverberation time is produced by the system using a maximum-likelihood parameter estimation procedure. A decision-making step then collects estimates of RT over a period of time and arrives at the most likely RT using an order-statistics filter. The method complements existing methods of RT estimation, being useful in situations where only passively received microphone signals are available.

II. THEORY

A model for blind estimation of reverberation time is presented. This is followed by an algorithm for implementation, and a decision-making strategy for selecting the estimate that best represents the reverberation time of listening rooms.

A widely used measure of the reverberation time is the T_{60} time first defined by Sabine (1922) and which is now a part of the ISO reverberation measurement procedure (ISO 3382, 1997). The T_{60} time measures the time taken for the sound level to drop 60 dB below the level at sound cessation. In practice, a decaying sound in a real environment reaches the ambient noise floor, thus limiting the dynamic range of the measured sound to values less than 60 dB, and so it is usually not possible to directly measure T_{60}. Instead, the decay rate is estimated by a ''linear least-squares regression of the measured decay curve from a level 5 dB below the initial level to 35 dB.'' [definition adopted from ISO 3382 (1997), p. 2]. If a 30-dB decay range cannot be measured, then a 20-dB range can be used. The T_{60} is simply the time taken to decrease by 60 dB from the initial level at the same decay rate given by the above measurements.

Before describing the model, we motivate the work with an example. The recorded response of a room to an impulsive sound source (a hand-clap) is shown in Fig. 1(a). As can be expected, there are strong early reflections followed by a decaying reverberant tail. If the early reflections are ignored, the decay rate of the tail can be estimated from the envelope. Figure 1(c) shows the measurement of T_{60} using the decay rate estimated from the -5- to -25-dB decay region. The procedure that was followed was that developed by Schroeder (1965a) described below.

We begin with a model for the diffusive or reverberant tail of sounds in a room. This refers to the dense reflections that follow the early reflections. All that can be said about them is that they are the result of multiple reflections, and appear in random order, with successive reflections being damped to a greater degree if they occur later in time. The assumption of randomness is crucial to the development of a statistical model. When a burst of white noise is radiated into a test enclosure, the phase and amplitudes of the normal modes are random in the instant preceding the cessation of the sound. Consequently, the decaying output of the enclosure following sound cessation will also be random, even if repeated trials were conducted with the same source and receiver geometry. Traditionally, and dating back to Sabine, the late decay envelope has been modeled as an exponential with

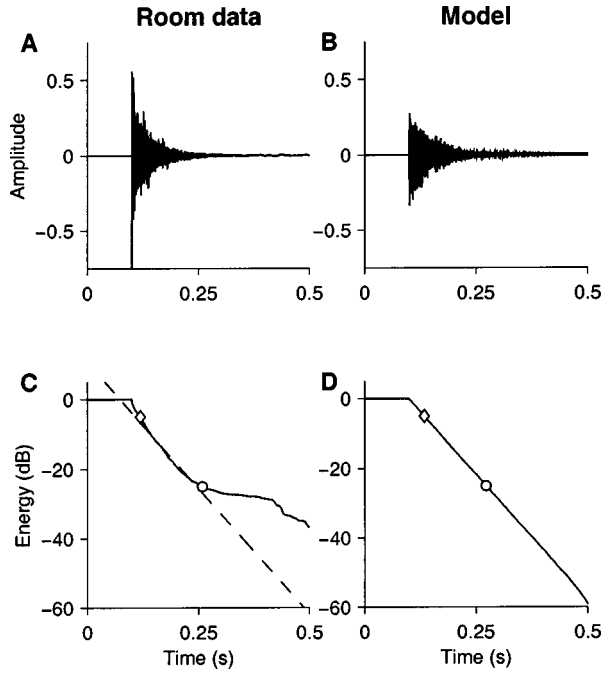

FIG. 1. Temporal decay of a hand-clap at $t=0.1$ s as recorded by a microphone (left column) and the model matching the reverberation (right column). (a) The recorded sound shows strong early reflections followed by a reverberant tail. Direct sound is excluded from the trace. (b) Model matching the reverberant tail shown in (a). Direct and early reflections are excluded. The model is a Gaussian white noise process damped by a decaying exponential, parametrized by the noise power σ and decay rate τ. (c) Decay rate estimated from Schroeder's backward integration method (Schroeder, 1965a) between -5 dB ($\diamond$) and -25 dB ($\circ$). Slope of linear fit (dashed line) yields $\tau=59$ ms ($T_{60}=0.4$ s). (d) Decay curve for model has identical slope everywhere following sound offset, and captures the most significant part of decay (-5 to -25 dB).

a single (deterministic) time-constant (hereafter referred to as decay rate). But, because the dense reflections are assumed to be uncorrelated, a convenient though highly simplified model is to consider the reverberant tail to be an exponentially damped uncorrelated noise sequence with Gaussian characteristics. The model does not include the direct sound or early reflections. The goal is to estimate the decay rate of the envelope.

A. Model of sound decay

We assume that the reverberant tail of a decaying sound y is the product of a fine structure x that is random process, and an envelope a that is deterministic. A central assumption is that x is a wideband process subject to rapid fluctuations, whereas the variations in a are over much longer time scales. Here, we will provide a statistical description of the reverberant tail with the goal of estimating the decay rate of the envelope.

Let the fine structure of the reverberant tail be denoted by a random sequence $x(n)$, $n\geqslant 0$, of independent and identically random variables drawn from the normal distribution $\mathcal{N}(0,\sigma)$. Further, for each n we define a deterministic sequence $a(n)>0$. The model for room decay then suggests that the observations y are specified by the 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 $\mathcal{N}(0,\sigma a(n))$. That is, the sequence $a(n)$ modulates the instantaneous power of the fine structure. For purposes of estimating the decay rate, we consider a finite sequence of observations, $n=0,\ldots,N-1$, where N will be referred to as the estimation interval, or estimation window length. For notational simplicity, denote the N-dimensional vectors of y and a by $\mathbf{y}$ and $\mathbf{a}$, respectively. Then the likelihood function of $\mathbf{y}$ (the joint probability density), parametrized by $\mathbf{a}$ and σ, is

$$L(\mathbf{y};\mathbf{a},\sigma)=\frac{1}{a(0)\cdots a(N-1)}\left(\frac{1}{2\pi\sigma^2}\right)^{N/2}\times\exp\left(-\frac{\Sigma_{n=0}^{N-1}(y(n)/a(n))^2}{2\sigma^2}\right), \tag{1}$$

where $\mathbf{a}$ and σ are the $(N+1)$ unknown parameters to be estimated from the observation $\mathbf{y}$. The likelihood function given by Eq. (1) is somewhat general, and, while it is possible to develop a procedure for estimating all $(N+1)$ parameters, suitable simplifications can be made when modeling sound decay in a room. Let a single decay rate τ describe the damping of the sound envelope during free decay. Then the sequence $a(n)$ is uniquely determined by

$$a(n)=\exp(-n/\tau). \tag{2}$$

Thus, the N-dimensional parameter a can be replaced by a scalar parameter a that is expressible in terms of τ and a single parameter $a=\exp(-1/\tau)$, so that

$$a(n)=a^n. \tag{3}$$

Introducing Eq. (3) into Eq. (1) yields

$$L(\mathbf{y};a,\sigma)=\left(\frac{1}{2\pi a^{(N-1)}\sigma^2}\right)^{N/2}\times\exp\left(-\frac{\Sigma_{n=0}^{N-1}a^{-2n}y(n)^2}{2\sigma^2}\right). \tag{4}$$

For a fixed observation window N and a sequence of observations $y(n)$, the likelihood function is parametrized solely by the decay rate a and the diffusive power σ.

The model is shown in Fig. 1(b) with parameters a and σ matched to the experimental hand-clap data shown in Fig. 1(a). Note that the model does not include the early reflections shown in panel (a). The Schroeder decay curve for the model is shown in Fig. 1(d) with a T_{60} time of 0.4 s in agreement with the measured T_{60}. The agreement between model and real T_{60} time motivates the search for an algorithm that can optimally estimate the two parameters.

B. Maximum-likelihood estimation

Given the likelihood function, the parameters a and σ can be estimated using a maximum-likelihood approach (Poor, 1994). First, we take the logarithm of Eq. (4) to obtain the log-likelihood function

$$\ln L(\mathbf{y};a,\sigma) = -\frac{N(N-1)}{2}\ln(a) - \frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1} a^{-2n}y(n)^2. \quad (5)$$

To find the maximum of $\ln(L)$, we differentiate the log-likelihood function Eq. (5) with respect to a to obtain the score function s_a (Poor, 1994):

$$s_a(a;\mathbf{y},\sigma) = \frac{\partial \ln L(\mathbf{y};a,\sigma)}{\partial a} \quad (6)$$

$$= -\frac{N(N-1)}{2a} + \frac{1}{a\sigma^2}\sum_{n=0}^{N-1} n a^{-2n}y(n)^2. \quad (7)$$

The log-likelihood function achieves an extremum when $\partial \ln L(\mathbf{y};a,\sigma)/\partial a = 0$; that is, when

$$-\frac{N(N-1)}{2a} + \frac{1}{a\sigma^2}\sum_{n=0}^{N-1} n a^{-2n}y(n)^2 = 0. \quad (8)$$

The zero of the score function provides a best estimate in the sense that $\mathbf{E}[s_a] = 0$.

Denote the zero of the score function s_a, and satisfying Eq. (8), by a^*. It can be shown that the second derivative $\partial^2 \ln L(\mathbf{y};a,\sigma)/\partial a^2|_{a=a^*} < 0$, i.e., the estimate a^* maximizes the log-likelihood function.

The diffusive power of the reverberant tail, or variance σ^2, can be estimated in a similar manner. Differentiating the log-likelihood function Eq. (5) with respect to σ, we have

$$s_\sigma(\sigma;\mathbf{y},a) = \frac{\partial \ln L(\mathbf{y};a,\sigma)}{\partial \sigma} \quad (9)$$

$$= -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_{n=0}^{N-1} a^{-2n}y(n)^2, \quad (10)$$

which achieves an extremum when

$$\sigma^2 = \frac{1}{N}\sum_{n=0}^{N-1} a^{-2n}y(n)^2. \quad (11)$$

As before, it can be shown that the $\mathbf{E}[s_\sigma] = 0$. Denote the zero of the score function s_σ, and satisfying Eq. (11), by σ^*. It can be shown that the second derivative $\partial^2 \ln L(\mathbf{y};a,\sigma)/\partial \sigma^2|_{\sigma=\sigma^*} < 0$, i.e., the estimate σ^* maximizes the log-likelihood function. Note that the maximum-likelihood equation given by Eq. (8) is a transcendental equation and cannot be inverted to solve directly for a^*, whereas the solution of Eq. (11) for σ^* is direct. Bounds on the variance of the estimates are presented in the Appendix.

C. Algorithm for estimating decay rate

Given an estimation window length and the sequence of observations $y(n)$ in the window, the zero of the score function Eq. (8) provides an estimate of a. The function is a transcendental equation that must be solved numerically using an iterative procedure. However, the estimate of σ can be obtained directly from Eq. (11). A two-step procedure was followed: (1) an approximate solution for a^* from Eq. (8) was obtained, and (2) the value of σ^* was updated from Eq. (11). The procedure was repeated, providing successively better approximations to a^* and σ^*, and so converging on the root of Eq. (8).

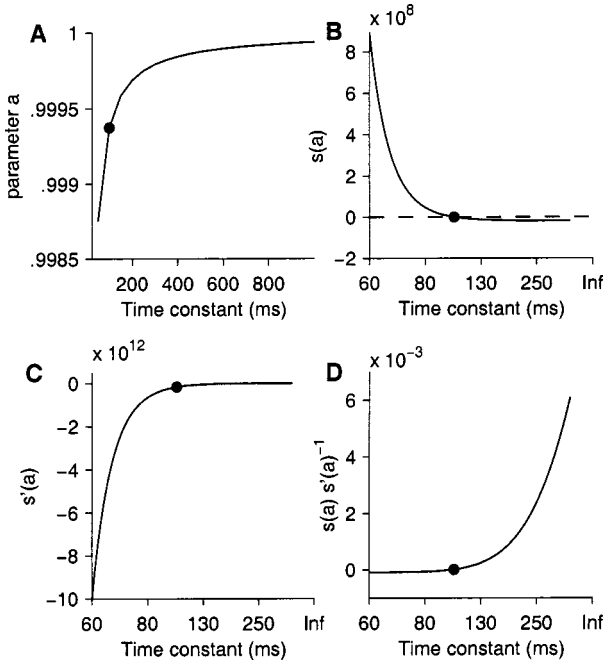

FIG. 2. Maximum-likelihood estimation (MLE) of room decay rate. (a) The decay rate of the exponential decay (τ, abscissa) is mapped to a parameter $a = \exp(-1/\tau)$ (ordinate) where τ is given in sampling periods. The function is monotone but highly compressive and maps $\tau \in [0,\infty)$ onto $a \in [0,1)$. Filled circle shows $\tau = 100$ ms ($a = 0.9994$). (b) Score function (derivative of log likelihood function) $s_a(a)$ (ordinate), decreases rapidly as a function of a [abscissa, marked in time constants using the map in (a)]. MLE of a is given by the root of $s(a)$ (filled circle). (c) The derivative $s_a'(a)$ as a function of a. At the root of s_a (filled circle), the derivative is negative. Note the nearly 8–12 orders of magnitude change in s_a and s_a' for commonly encountered values of τ. (d) The ratio $s_a(a)/s_a'(a)$ (ordinate) as a function of a is the incremental step size of the Newton–Raphson procedure for finding the root of Eq. (8). It provides an estimate of the convergence properties of the root-finding algorithm. Sampling frequency was 16 kHz, and the log-likelihood function was calculated assuming a 400-ms window.

Here we address the strategy for extracting the root in the smallest number of iterative steps. To gain an understanding of the root-solving procedure, we consider the example shown in Fig. 2. The function $a = \exp(-1/\tau)$ maps the room decay rate τ one-to-one and onto a as shown in Fig. 2(a). For instance, consider a room decay rate of 0.1 s and a sampling rate of 16 kHz. Then the decay rate is 1600 samples, and so $a = 0.9994$ (filled circle). The significance of the number becomes clear if we consider that when $\tau = 0.03$ s, then $a = 0.9979$, whereas for $\tau = \infty$, $a = 1$. Hence the geometric ratio is highly compressive and values of a for real environments are likely to be close to 1. Thus, the advantage of estimating a rather than τ is due to the bounded nature of a. The score function s_a from Eq. (7), on the other hand, has a wide range [about eight orders of magnitude, see Fig. 2(b)] and is zero at the room decay rate (filled circle). The gradient of the score function ds_a/da shown in Fig. 2(c) also demonstrates a wide range, but takes a negative value at the zero of s_a.

Thus, if we start with an initial value of $a_0^* < a$, the root-solving strategy must descend the gradient sufficiently

rapidly. The standard method for solving this kind of nonlinear equation, where an explicit form for the gradient is available, is the Newton–Raphson method which offers second-order convergence (Press *et al.*, 1992). The order of convergence can be assessed from $s_a(ds_a/da)^{-1}$ which is the incremental step size Δa in the iterative procedure [Fig. 2(d)]. For example, with true value of $\tau=100$ ms, Δa at intermediate values in the iteration can be as small as 10^{-6} when $a=0.9993$ ($\tau=90$ ms) or $a=0.9995$ ($\tau=120$ ms). This corresponds to an incremental improvement of about 0.01 ms for every iteration, thus providing slow convergence if the initial value is far from the zero. On the other hand, the bisection method (Press *et al.*, 1992) guarantees rapid gradient descent but works poorly in regions where the gradient changes relatively slowly (such as near the true value of a). Furthermore, it guarantees only first-order convergence.

However, the specific structure of the root-solving problem can be exploited because the behavior of s_a is known. Here, both methods were used to obtain rapid convergence to the root. First, the root was bisected until the zero was bracketed, after which the Newton–Raphson method was applied to polish the root. For the example shown, the root bracketing was accomplished in about eight steps and the root polishing in two to four steps. In contrast, with the same initial conditions, the Newton–Raphson method took about 500 steps to converge. Taken together, the analysis presented here suggests that the estimation procedure is feasible and does not lead to significant errors although values of a for real rooms are close to 1, and the score function and its derivative vary over many orders of magnitude. While other root-solving procedures are possible, such as iterative gradient optimization, these are not dealt with here.

D. Strategy for assigning the correct decay rate from the estimates

The theory presented in the preceding section provides one estimate of a and σ in a given time frame of N samples. By advancing the frame as the signal evolves in time, a series of estimates a_k^* will be obtained, where k is the time frame. Some of these estimates will be obtained during a free decay following the offset of a sound segment (correct estimations), whereas some will be obtained when the sound is ongoing (incorrect estimations due to model failure). Thus, a strategy is required for selecting only those estimates that correctly represent regions of free decay and hence the real room decay rate. This requires a decision-making strategy that examines the distribution of the estimates after a sufficient number of frames have been processed, and makes a decision regarding the true value of the room decay rate.

In a blind estimation procedure the input is unknown, and so the model will fail when (1) an estimate is obtained in a frame that is not occurring during a free decay. This includes regions where there is sound onset or sound is ongoing. In these periods, the MLE scheme can provide widely fluctuating or implausible estimates due to model failure. (2) The model will also fail during a region of free decay initiated by a sound with a gradual rather than rapid offset. In this case, the offset decay of the sound will be convolved with the room response, prolonging the sound even further and, so, the estimated decay rate will be larger than the real room decay rate. Gradual offsets occur in many natural sounds, such as terminating vowels in speech. We address both issues here and provide a strategy for selecting the correct room time constant.

In the first case where the estimation frames do not fall within a region of free decay, many of the time frames will provide estimates of a close to unity (i.e., infinite τ) or implausible values. On the other hand, the estimates will accurately track the true value when a free decay occurs. Intuitively, a strategy for selecting a from the sequence a_k^* is guided by the following observation: the damping of sound in a room cannot occur at a rate *faster* than the free decay, and thus all estimates a^* must attain the true value of a as a lower bound. The bound is achieved only when a sound terminates abruptly, upon which the model conditions will be satisfied, and the estimator will track the true value of the decay rate.

Although it seems intuitive to set $a=\min\{a_k^*\}$, it should be recognized that even during a free decay the estimate is inherently variable (due to the underlying stochastic process), and so selecting the minimum is likely to underestimate a.

A robust strategy would be to select a threshold value of a^* such that the left tail of the probability density function of a^*, $p(a^*)$, occupies a prespecified percentile value γ. This can be implemented using an order statistics filter specified by

$$a=\arg\left\{P(x)=\gamma:P(x)=\int_0^x p(a^*)da^*\right\}. \tag{12}$$

For a unimodal symmetric distribution with $\gamma=0.5$ the filter will track the peak value, i.e., the median. Order-statistics filters play an important role in robust estimation, especially when data is contaminated with outliers (Pitas and Venetsanopoulos, 1992), as is the case here. It should be noted that for γ values approaching 0, the filter Eq. (12) performs like the minimum filter $a=\min\{a_k^*\}$ suggested above.

In the second case described above, where the sound offset is gradual, $p(a^*)$ is likely to be multimodal because sound offsets (such as terminating phonemes in speech) will have varying rates of decay, and their presence will give rise to multiple peaks. The strategy then is to select the first dominant peak in $p(a^*)$ when a^* is increasing from zero (i.e., left most peak), that is,

$$a=\min\arg\{dp(a^*)/da^*=0\}, \tag{13}$$

where the minimum is taken over all zeros of the equation. If the histogram is unimodal but asymmetric, the filter tracks the mode and resembles the order-statistics filter.

In connected speech, where peaks cannot be clearly discriminated or the distribution is multi-modal, Eq. (12) can be employed by choosing a value of γ based on the statistics of gap durations. For instance, if gaps constitute approximately 10% of total duration, then $\gamma=0.1$ would be a reasonable choice. A judicious choice of γ can result in the filter per-

forming like an edge detector, because it captures the transition from larger to smaller values of the time-evolving sequence a_k^*.

The decision strategies, as depicted in Eqs. (12) and (13), were used to validate the model in simulated and real environments (see Sec. IV).

III. EXPERIMENTAL METHODS

In addition to simulations, the MLE approach was validated with real room data. The experimental methods and data analysis procedures are described in the following sections.

A. Sound recordings

To validate the MLE method, sound recordings were made in several rooms, building corridors and an auditorium, with the aim of determining their reverberation times. Sound stimuli that were used included 18-tap maximum length sequences (period length of $2^{18}-1$), clicks (100 μs), handclaps, word utterances (International Phonetic Assoc., 1999), and connected speech from the Connected Speech Test (CST) corpus (Cox *et al.*, 1987). Recordings were made using a Sennheiser MK-II omni-directional microphone (frequency response 100–20 000 Hz). Microphone cables (Sennheiser KA 100 S-60) were connected to the XLR input of a portable PC-based sound recording device (Sound Devices USBPre 1.5). The recorder transmitted data sampled at 44.1 kHz to a laptop computer (Compaq Presario 1700, running Microsoft Windows XP) via a USB link. The sound stimuli, stored as single-channel presampled (44.1 kHz) WAV files, were played through the headphone output of the laptop, amplified by a power amplifier (ADCOM GFA-535II) and presented through a loudspeaker (Analog and Digital Systems Inc., ADS L200e). Data acquisition and test material playback were controlled by a custom-written script in MATLAB (The MathWorks, Inc.) using the Sound PC Toolbox (Torsten Marquardt).

B. Measurement of T_{60} time using Schroeder's method

To validate the estimation procedure, experimentally recorded data from real listening environments were processed using the MLE procedure and compared to results obtained from a widely used method of Schroeder (1965a). Experimentally, RT is determined from decay curves obtained by radiating sound into the test enclosure. The sound source is switched on, and when the received sound level reaches a steady state, it is switched off. The decay curve is the received signal following the cessation of the sound source, according to the Interrupted Noise Method (ISO 3382, 1997). When the excitation signal is a noise source, the decay curve will be different from trial to trial due to random fluctuations in the signal, even when the experimental conditions are unchanged. This is in part due to the random phase and amplitudes of the normal modes at the moment the excitation signal is turned off. Prior to Schroeder's method, fluctuations in RT estimates were minimized by averaging the RTs obtained from many decay curves. Schroeder (1965a) developed an alternate method that, in a single measurement, yields the average decay curve of infinitely many interrupted noise experiments. Thus, Schroeder's method eliminates the averaging procedure.

Following Schroeder (1965a), let $n(t)$ be a stationary white noise source with power σ^2 per unit frequency, and $r(t)$ be the impulse response of the system consisting of the receiver, transmitter, and the enclosure. Then a single realization of the decay curve $s(t)$ from the interrupted noise experiment is given by

$$s(t)=\int_{-\infty}^{0} n(\tau)r(t-\tau)d\tau, \tag{14}$$

where the noise is assumed to be switched off at $t=0$, and the lower limit is meant to signify that sufficient time elapsed for the sound level to reach a steady state in the enclosure before it was switched off. The reverberation time is obtained from the decay curve $s(t)$ (see below).

In practice, the experiment was repeated to obtain a large number of decay curves, and RTs from these curves were averaged. Schroeder used Eq. (14) to establish a relationship between the mean squared average of the decay curve $s(t)$ and the impulse response of the enclosure $r(t)$, namely,

$$\mathbf{E}[s^2(t)]=\sigma^2\int_{t}^{\infty} r^2(\tau)d\tau. \tag{15}$$

While the ensemble average on the left-hand side requires averaging over many trials, the right-hand side requires only a single measurement, as it is the impulse response of the enclosure plus receiver and transmitter.

Schroeder's method, called the Integrated Impulse Response Method (or sometimes, Backward Integration Method), can be applied to a single broadband channel (say an impulsive sound covering a broad range of frequencies) or to a narrow-band channel consisting of a filtered impulse (such as a pistol shot). The only requirement is that the power spectrum of the excitation signal [in Schroeder's method, right-side of Eq. (15)] should be identical to the power spectrum of the noise burst [in the noise decay method, left-side of Eq. (15)].[1]

The recorded data were filtered offline in ISO one-third octave bands (21 bands with center frequencies ranging from 100 to 10 000 Hz) using a fourth-order Type II Chebyshev band-pass filter with stopband ripple 20-dB down. The output from each channel was processed by the MLE procedure and Schroeder's method using Eq. (15). For broadband estimation, the microphone output was processed directly using the two methods.

Due to the limited dynamic range of sounds in real environments, Schroeder's method requires the specification of a decay range. The decay ranges normally used are from -5 to -25 dB (20-dB range), and from -5 to -35 dB (30-dB range). The decay curves in each range were fitted to a regression line using a nonlinear least squares fitting function (function nonlinsq provided by MATLAB). The fitted function was of the form Aa_d^n, where A is a constant, n is the sample number within the decay window, and a_d is the geometric

ratio related to the decay rate τ_d of the integrated impulse response curve by $a_d = \exp(-1/\tau_d)$. This is in contrast to the model depicted in Eq. (2) which assumes an exponentially decaying envelope with decay rate τ, whereas Schroeder's decay curve is obtained by squaring the signal. Hence, $\tau_d = \tau/2$. Two estimates of the decay rate were obtained from decay curves fitted to the −5- to −25-dB and −5- to −35-dB drop-offs. For each fit, the line was extrapolated to obtain T_{60} time (in seconds) using the expression

$$T_{60} = \frac{6}{\log_{10}(e^{-1})\log_e(a_d)} = \frac{-6\tau_d}{\log_{10}(e^{-1})} = 13.82\tau_d. \quad (16)$$

The same procedure was followed for determining the decay rate from broadband signals. It should be noted that the MLE procedure does not require the specification of a decay range, but only the specification of the estimation window length; thus, only one estimate per band is obtained.

C. Verification of MLE procedure with ideal stimuli

Microphone data were processed using the MLE procedure to obtain a running estimate of the decay rate. For model verification, estimation was performed on (1) the segment following the cessation of a maximum-length sequence or a hand-clap, and (2) the entire run of a string of isolated word utterances. These were considered ideal stimuli, because they fulfilled the model assumptions of free decay or possessed long gaps between sound segments. The estimates were binned for each run and a histogram was produced. The histogram was examined for peaks, and the decay rate was selected using the order-statistics filter Eq. (13) if there were multiple peaks, or Eq. (12) if the histogram was unimodal. The estimate $\hat{a}$ so obtained was used to calculate T_{60} (in seconds) using the formula

$$T_{60} = \frac{3}{\log_{10}(e^{-1})\log_e(\hat{a})} = \frac{-3\tau}{\log_{10}(e^{-1})} = 6.91\tau. \quad (17)$$

In theory, the T_{60} expressions given by Eqs. (16) and (17) are identical due to the relationship between τ and τ_d. However, the calculated values may differ, and this can be ascribed to either model inadequacies or discrepancies in measurement and analysis.

D. Verification of MLE procedure for speech

The performance of the MLE was also verified using connected speech played back in a circular building foyer (6-m diameter). Test materials were connected sentences from the CST corpus. Estimates from nonoverlapping 1-s intervals were binned to yield a histogram, and the first dominant peak from the left of the histogram was selected to determine the room decay rate. The procedure for calculating T_{60} time followed Eq. (17).

IV. RESULTS

The estimation procedure was applied to a variety of data sets, including simulated data and real room responses. To illustrate the methods and identify the strengths and deficiencies of the estimation procedure, we first consider simulated data sets. Subsequently we will provide results for real data that validate the room decay rate estimates, and compare these to results from Schroeder's method.

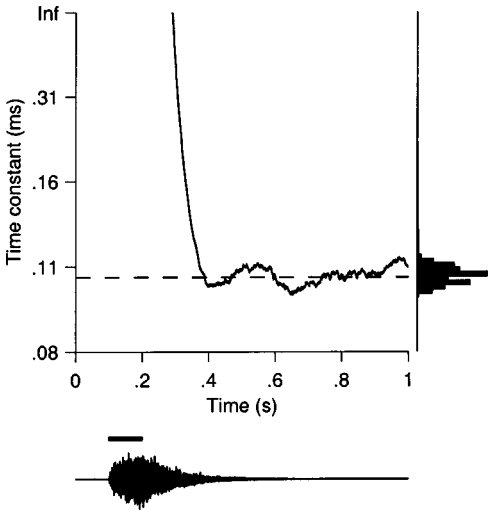

FIG. 3. Illustration of procedure for continuous estimation of decay rate. A burst of white noise was applied at time $t = 0.1$ s (black bar, bottom trace, 100-ms duration). Simulated room output (bottom trace) shows the buildup and decay of sound in the room. A running estimate of the parameter a in 200-ms windows is shown in the graph (ordinate, a converted to decay rate in seconds). The true value of decay rate (100 ms) is shown as horizontal dashed line. The estimation window was advanced by one sample to obtain the trace, with each point at time t being the estimate in the window $(t - 0.2, t]$. During the buildup and ongoing phase of the sound, estimated a sometimes exceeded 1 (i.e., negative values of τ). These were discarded and are not shown. As the window moved into the region of sound decay ($t > 0.3$ s), the estimates converged to the correct value. A histogram of the estimated decay rate is shown to the right of the trace. An order-statistics filter, such as the mode of the histogram, can be used to extract the room decay rate. Sampling rate was 16 kHz.

A. Broadband white noise bursts in simulated rooms

A 100-ms burst of broadband white noise (8-kHz bandwidth) was convolved with the impulse response of a simulated room having a decay rate $\tau = 100$ ms (Fig. 3). Room output (bottom trace of Fig. 3) shows the characteristic rise and decay of sound following onset and offset of noise burst (horizontal bar). The graph shows the running estimate of decay rate obtained in a 200-ms time window by advancing every sample. Time frames up to about $t = 0.3$ s are not regions of free decay, and so the estimator tended to produce values of $a > 1$. When this was observed in the root-bracketing step of the estimate, the root-solving procedure was aborted. Thus all estimates of a were bounded above by 1. It can be seen that when the window crosses into the region of free decay, the estimator output stabilizes at the true value (horizontal dashed line). A histogram of the decay rate estimates (right axis) was input to the order statistics filter Eq. (12) with $\gamma = 0.5$. The reported decay rate from the filter was $\tau = 101$ ms.

For comparison, the procedure was repeated with the simulated noise burst input (i.e., before it was convolved with the room impulse response) to mimic anechoic conditions. The histogram of a^* demonstrated a strong peak at $a = 1$ $(\tau = \infty)$ (not shown). This showed that in the absence of

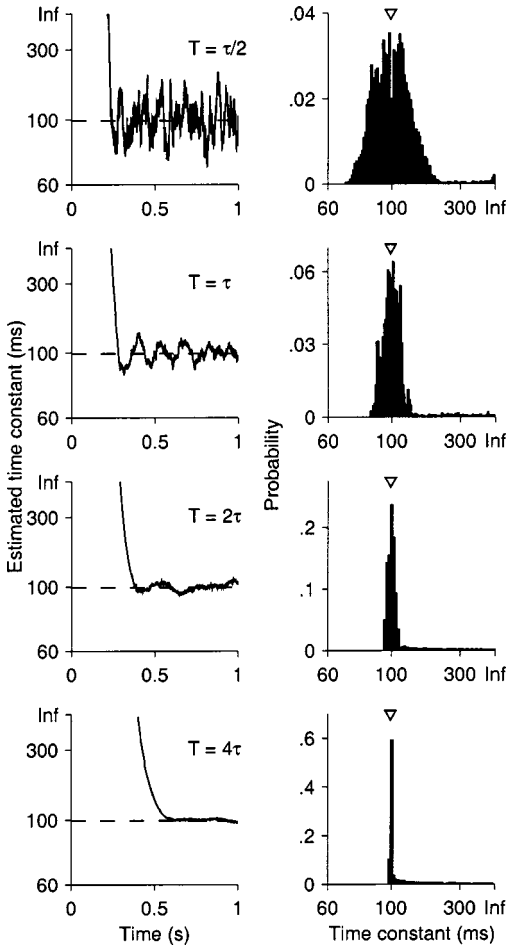

FIG. 4. Effect of estimation window length on the variance of the estimate. The simulation shown in Fig. 3 was repeated for windows of duration 0.5τ, τ, 2τ, and 4τ (top to bottom), where $\tau=100$ ms is the true value of the room decay rate. The left column shows the running estimate of parameter a (ordinate, shown as decay rate in ms) as a function of time (abscissa). The right column shows the histogram of the estimates. The variance of the estimate decreases with increasing window length (arrowheads mark true value of τ).

reverberation, as in an anechoic environment or open space, histograms showing strong peaks at $a=1$ are to be expected.

B. Effect of window length on estimation

A parameter that is critical for estimation performance is the window length N specified in Eq. (8). Small window lengths are expected to increase the variance of the estimate, as also indicated by the Crámer–Rao lower bound [Eq. (21)]. To test the effect of window length a burst of white noise (100-ms duration) was convolved with a simulated room impulse response ($\tau=100$ ms), and the estimator tracked the decay curve using four different window lengths. The results are shown in Fig. 4. As window length increased from 0.5τ to 4τ, the MLE procedure gave improved estimates. Further, for all four window lengths, there was no bias in the estimates of the peak position. We concluded that increasing window length reduced the variability in the estimates, and did not introduce significant bias.

Although it is desirable to have long window lengths, in practice this is limited by the duration and occurrence of gaps between sound segments. Ideally the filter length should be of the order of τ or longer, but if the gaps are short, then increasing the filter length beyond the mean gap will produce undesirable effects where the next sound segment creeps into the window. Thus, the window length should not be less than one-half or one-third of τ, but the upper limit is dictated by the mean duration of gaps.

C. Speech sounds in simulated room

The examples considered above illustrated the performance of the algorithm when the input was broadband white noise. To be applicable in realistic conditions, the algorithm must perform in a variety of conditions and with different signal types. Speech represents an example where the algorithm is expected to perform poorly, because it is nonstationary and non-Gaussian. Further, the offset transients in speech sounds (including plosives) have a natural decay rate (not to be confused with the room decay rate) that can vary from 5–40 ms.[2] Thus, estimation of room decay rate with speech presents a challenge to the algorithm. We took a sequence of 15 distinct and isolated American-English words recorded in an anechoic environment at a sampling rate of 20 kHz (International Phonetic Assoc., 1999). These included 11 consonant–vowel–consonant words (/p,b,g/V/d/, e.g., ''bed''), and four consonant–vowel words (/b/V/, e.g., ''bay'') separated by a mean interval of 200 ms. These were convolved with a simulated room impulse response having decay rate $\tau=100$ ms. The task of the estimator was to track the decays for the entire duration of the sequence (approximately 11.4 s). The control condition was the clean input (i.e., anechoic). The results are shown in Fig. 5. Four different filter lengths were used as in Fig. 4. For the control condition (left column) no reliable estimates were produced for the smallest three windows (top three panels) because the histogram peaked at values of τ approaching ∞. For the simulated room response (right column), the peak shifted towards the true value of τ, with the best estimates being obtained for the largest window size of 4τ (right column, bottom row). In all the histograms the peak was located at about 115 ms (arrow). This estimate deviated from the real decay rate of 100 ms due to the lack of sharp transients in the clean speech. A gradual sound offset tends to prolong the reverberated sound even further. This can be seen in the ''anechoic'' control condition where a small peak is noticeable when window size is 4τ (bottom panel, left column). The peak occurs around 60 ms, and corresponds to the gradual offsets of speech sounds. Thus, this introduces a bias in the estimates under reverberant conditions.

The results of the preceding sections demonstrate the importance of selection of a suitable estimation window length. The choice of window length determines the variability of the estimates, and is critical to obtaining a histogram with a clearly resolved peak at the true value of the room decay rate. However, the effect of variability on the order-statistics filter is difficult to determine as the filtering operation is nonlinear. Further, bimodal or multimodal histograms may be obtained if there is fluctuating background noise or if the sound segments have an intrinsic offset decay rate (as shown in Fig. 5).

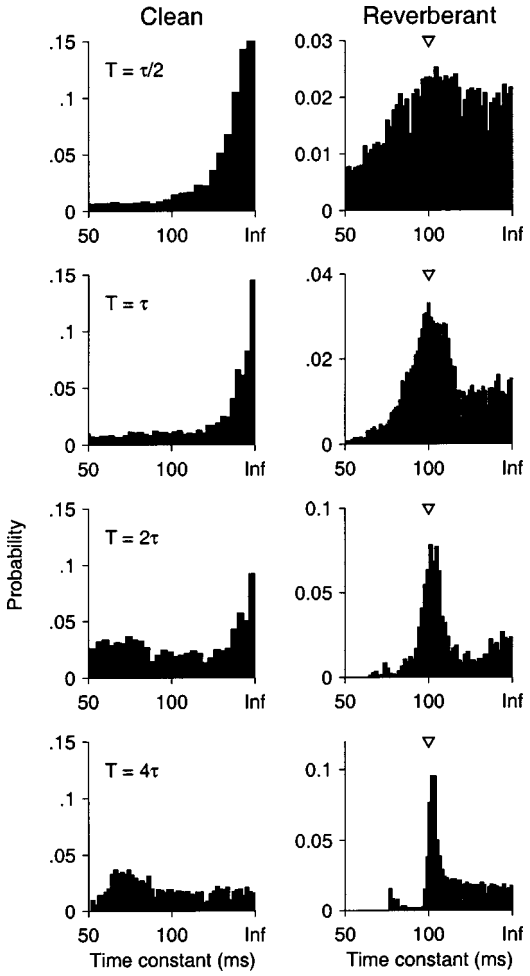

FIG. 5. Estimation of room decay rate from speech. Fifteen words recorded in an anechoic (clean) environment (200-ms interword spacing) were convolved with a room model (τ=100 ms). Histograms of decay rates were estimated from clean (left column) and simulated reverberant responses (right column), and are shown for window durations 0.5τ, τ, 2τ, and 4τ (top to bottom). The histogram for clean speech served as a control. Description follows Fig. 4. Estimation from reverberant speech produces a clearly defined peak, especially for the longer window lengths, albeit with a small bias (right column, 2τ and 4τ). The bias can be attributed to the gradually decaying offsets inherent in speech so that the resultant decay is a convolution of speech offset and the room response. For the control condition (left column), the offset decay is visible only in the bottom two rows where the histogram exhibits a broad bump between 50 and 100 ms. The 15 words included 11 /p,b,g/V/d/ and 4 /b/V/ sampled at 20 kHz.

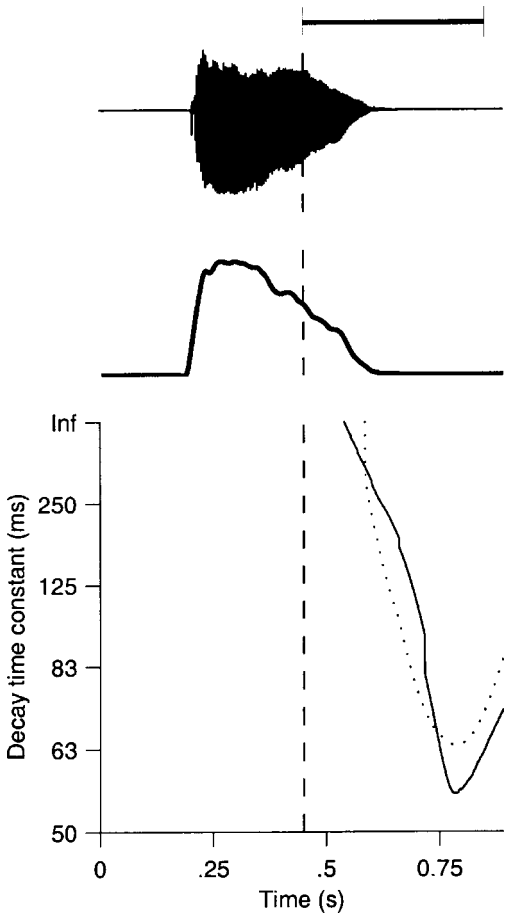

FIG. 6. Illustration of decay rate estimation when a terminating phoneme is encountered. The word ''bough'' recorded under anechoic (clean) conditions (top row) has a gradually decaying offset. The envelope was extracted by filtering the absolute value of the analytic signal (second row, heavy outline), and its decay rate was estimated for the segment following the dashed line using two methods (bottom row). Overlapping segments (duration given by bar, with step size indicated by the thickness of the vertical end) were converted to a decibel scale and the decay rate obtained by a least squares fit to a straight line (dotted trace). The same segments were analyzed using the MLE algorithm to obtain a second estimate of the decay rate (solid trace). While the estimators provide somewhat different results, they are in qualitative agreement. Both methods suggest that the fastest decay rate is in the range of 50–70 ms (see also Fig. 5).

D. The effect of gradual offsets in speech sounds on decay rate estimation

The preceding section introduced the problem of estimating the room decay rate when the input signal exhibited varying offset decay time courses. Here we examine in greater detail the performance of the estimator with input comprising a single word (/b/V/, ''bough''). The word was recorded under anechoic conditions and presented to the estimator without modification so that the effect of the vowel offset could be determined. The results are shown in Fig. 6. The terminating vowel has a gradually decaying offset (top panel). Estimation of the offset decay was performed from $t=0.45$ s (vertical dashed line) using two procedures. First, the envelope was extracted from the analytic signal via a Hilbert transform, windowed, and filtered to eliminate frequency components above 100 Hz. The envelope is shown in the middle panel (heavy outline). The envelope was then squared and transformed to a decibel scale, and the decay rate was estimated in windows of duration 0.4 s (horizontal bar), using a least squares fit to a straight line. Successive estimates were obtained by sliding the window forward in steps of one sample. Note that the time at which an estimate is reported for any given window is the end point of the window. The estimate for the window indicated by the horizontal bar, for instance, is plotted at time $t=0.85$ s. A curve of the estimated decay rates was thus obtained (dotted curve, bottom panel). The envelope-based method employed here is similar to the method of Lebart *et al.* (2001), except that they performed a one-time RT estimation over the entire decay period using linear regression. The MLE procedure was applied to the same segments and produced an independent estimate of the decay rate (solid line, bottom panel). While the estimates differ somewhat, they are in qualitative agree-

ment. Both procedures indicate that the terminating vowel had a time-dependent decay rate, and the greatest rate was between 50 and 70 ms.

The results confirm the presence of the peak in Fig. 5 (left column, bottom panel), although the histogram shown in Fig. 5 was obtained for a sequence of 15 words. The analysis shown in Fig. 6 also indicates the reason for estimation bias under reverberant conditions using speech samples. The offset decays present in clean speech segments will be convolved with the room impulse response, and the estimated decay rates will consequently be greater than the room decay rate. Taken together, the results from Figs. 4–6 suggest that the factors responsible for estimation performance are the presence of adequate numbers of gaps, sharp offset transients, and estimation window length.

E. Validation of method

The above results demonstrate that estimation of decay rate is possible for a variety of sounds including impulses, noise bursts, and speech. Although we have shown that a reasonable agreement exists with a nonlinear least squares fit to the data (Fig. 6), a more careful evaluation is necessary to determine the conditions under which the MLE procedure is likely to provide accurate estimates. Here we establish that the estimated decay rates are comparable to decay rates obtained from the method by Schroeder (1965a). Furthermore, any data collected must be under sufficiently realistic conditions where there is background noise and where the testing sound is not subject to experimental control. A comparison of MLE performance with the standard method in real environments will therefore establish the utility of the method.

We compared the estimates using the method by Schroeder (1965a) in both single-channel (i.e., the broadband signal), and multi-channel frameworks (i.e., narrow-band signals occupying ISO one-third octave bands). Schroeder's method requires a fitting procedure to estimate the decay rate in a preselected decay range (either 20 or 30 dB below a reference level of −5 dB, see Sec. III). The MLE procedure does not require the specification of such a range.

To determine whether the two methods provide the same RT value, estimations were performed on a simulated room decay curve with RT=0.5 s (Fig. 7). Broadband and one-third octave band estimates were obtained using the MLE method (circle) and Schroeder's method (20 dB: lozenge, 30 dB: square). Figure 7(a) shows the mean value of RT as a function of center frequency of the one-third octave bands (open symbols) and the broadband estimate (filled symbols near *y* axis). range) averaged over 100 trials. The broadband estimates were 0.504 s (MLE) and 0.5 s (Schroeder's method) for both 20- and 30-dB decay ranges. While the MLE estimate was significantly different from Schroeder's method ($p<0.0001$, Wilcoxon rank sum test), the discrepancy was less than 1%. The one-third band MLE estimates in most cases were somewhat higher than the Schroeder estimates by about 0.5% (mean RT over all bands were, MLE: 0.505, Schroeder's method: 0.502 s for 20 dB and 0.501 s for 30 dB). However, the estimates were not significantly different, except for one estimate obtained from the 30-dB decay curve in the band centered at 8 kHz. The most noticeable

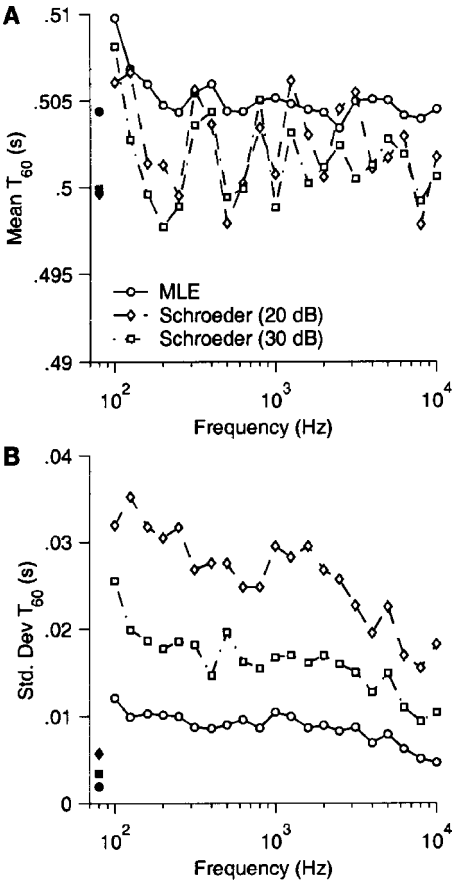

FIG. 7. Comparison of RT estimates obtained from MLE method and Schroeder's method. (a) Mean RT (ordinate) in one-third octave bands (abscissa) averaged over 100 independent trials of a simulated decay curve (RT=0.5 s). RT estimates were obtained using the MLE procedure (circles), and Schroeder's method in 20-dB decay range (lozenge), and 30-dB decay range (square). The filled symbols are broad-band estimates. (b) Standard deviation of the RT for broadband and one-third octave bands over 100 trials. Symbols follow (a).

difference between the two methods was in the variability of the estimates as measured by the standard deviation over the trials [Fig. 7(b)]. The MLE method demonstrated lower SD across trials than Schroeder's method, by factors of 2 (for the 20-dB curve) and 3 (for the 30-dB curve). Further, MLE estimates were similar across one-third octave bands at frequencies above 200 Hz [Fig. 7(a)], whereas estimates from Schroeder's method exhibited greater variability. The results establish that the MLE method and Schroeder's method are in good agreement when tested on model data. While the MLE method may overestimate the RT when using broadband signals (although this is no more than 1%), the narrow-band estimates are comparable to those obtained from Schroeder's method, are consistent over a wide range of frequencies, and subject to less variability.

We first report on the comparison between the methods using a hand-clap in a small office (8×3×3 m). Subsequently we will summarize results obtained in rooms of different sizes. Figures 8(a) and (b) depict a hand-clap event and its spectrogram, respectively. The data in panel (a) is the same as shown in Fig. 1(a), except that Fig. 8(a) also includes the direct sound. The rms noise level in the room was 50 dBA SPL, and the peak sound pressure level resulting

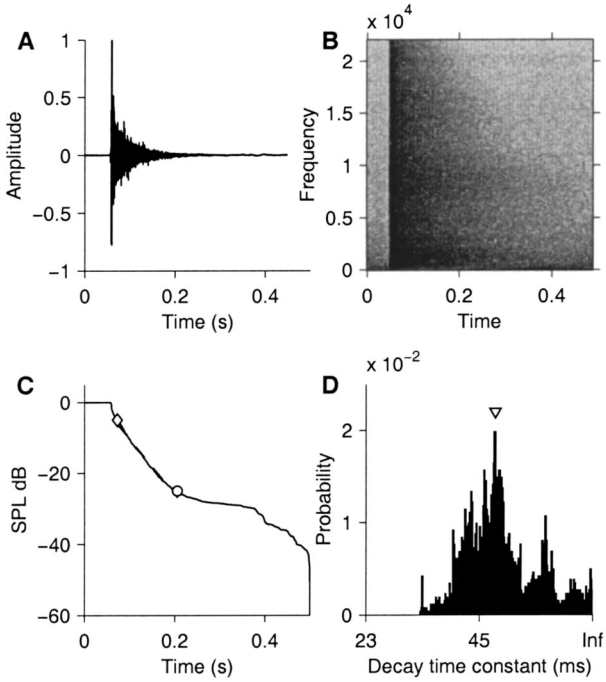

FIG. 8. Estimation of decay rate from real room data. (a) The room response to a hand-clap [same as Fig. 1(a) but includes the direct sound]. (b) Spectrogram of the hand-clap demonstrates a sharp broadband onset transient and the decay as a function of frequency. (c) The decay rate was estimated using Schroeder's backward impulse integration method in the −5-dB (lozenge) to −25-dB (circle) range, followed by a least-squares fit to a straight line to obtain the decay rate (τ=56 ms, T_{60}=0.39 s). (d) Histogram of decay rate obtained from signal shown in (a) using MLE. The median value of the histogram (arrow) is τ=53 ms, T_{60}=0.37.

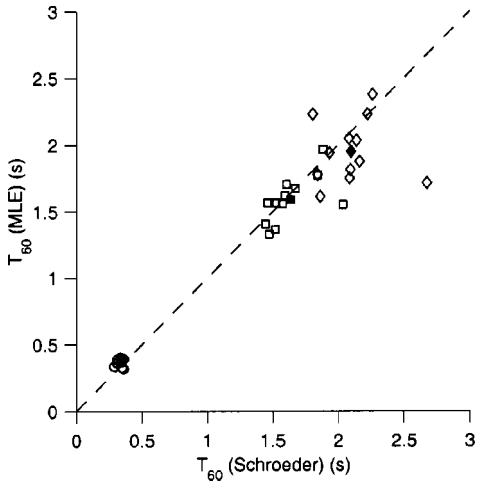

FIG. 9. Comparison of Schroeder's method and the MLE procedure for T_{60} times obtained in one-third octave bands. Three environments were tested: a moderately reverberant environment (circles; the environment is the same as shown in Fig. 8), a highly reverberant circular foyer (squares), and a highly reverberant enclosed cafeteria (diamonds). In each environment, a single hand-clap was filtered using a bank of ISO one-third octave band-pass filters with center frequencies exceeding 1 kHz. The ordinate shows the best estimates obtained from the MLE procedure for each band, and the abscissa shows the T_{60} times obtained from Schroeder's method. Averages over all bands for each environment are shown as filled symbols. The diagonal dashed line (with unity slope) is shown for reference, and points lying close to this line suggest good agreement between the two procedures.

from the hand-clap was 85 dBA SPL. The decay curve obtained using Schroeder's method is shown in Fig. 8(c), normalized so that the peak SPL was 0 dB. This is the broadband curve obtained by integrating the recorded microphone signal. A straight-line fit to the 20-dB drop-off point (circle) from a reference level of −5 dB (lozenge) yielded τ=56 ms (T_{60}=0.39 s). The discrepancy between this value and that presented in Fig. 1 (τ=59 ms) was due to the inclusion of the direct sound in Fig. 8. The windows over which the 20-dB drop-off was computed were not identical for the two cases. The data were run through the MLE procedure and a histogram of estimates was obtained, and the decay rate was calculated from the peak of the histogram using Eq. (12). This gave an estimate τ=53 ms (T_{60}=0.37 s), which is in good agreement with the estimate obtained using Schroeder's method. Note that the estimates reported in this work are based on a single trial. The normal practice is to average over large numbers of trials. However, our goal is to develop an online estimation procedure, and so we felt that it would be more realistic to use a single trial.

To test a range of room RTs, ISO one-third octave band analysis (exceeding 1 kHz center frequency) was performed in three environments. These were (1) the moderately reverberant room described above (Fig. 8), (2) a highly reverberant circular foyer, and (3) a highly reverberant enclosed cafeteria. In all cases, the signal was a hand-clap generated at a distance of 2 m from the recording microphone (peak value 90 dB SPL). Output from the band-pass filters were analyzed using the MLE procedure, and the τ value for each band was obtained from the histogram by selecting the dominant peak. For Schroeder's method, a 20-dB decay range was used. Figure 9 shows the T_{60} estimates from Schroeder's method (abscissa) versus the MLE estimates (ordinate) for each ISO one-third band (open symbols), and the average over these bands (closed symbols).

Figure 9 shows that the variability of estimates for highly reverberant environments increases with increasing mean RT for both methods. However, the two methods are in good agreement, especially in the high-frequency bands (the single outlier falling below the diagonal in Fig. 9 is the lowest center frequency used in the analysis, namely 1 kHz). The agreement between the methods is best when the T_{60} values are averaged over all bands (filled symbols), as is usually reported in the literature.

A more extensive test to determine the variability in estimates across different environments, and between bands, was performed in 12 environments, including small office rooms, an auditorium, large conference rooms, corridors, and building foyers. The data were analyzed as in Fig. 9 and are shown in Fig. 10(a). In comparison with Schroeder's method, the MLE procedure consistently overestimated T_{60} in low to moderately reverberant environments ($T_{60}<0.3$ s) whereas it underestimated the reverberation time for more reverberant environments ($T_{60}>1.3$ s). There was a good agreement between the two methods for intermediate ranges. The average T_{60} over all bands (filled squares) were, however, in good agreement. Broadband estimates were made using the same procedures but without band-pass filtering of recorded signals. These are shown in Fig. 10(b). The trend in the estimates was similar to that observed with narrow-band

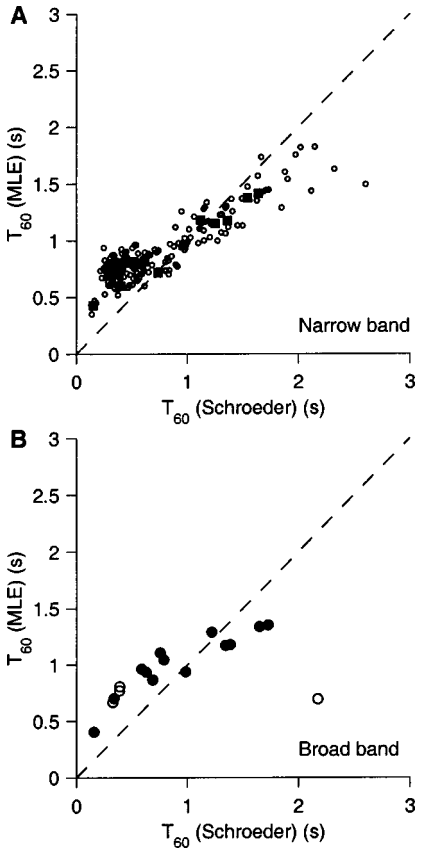

FIG. 10. Reverberation-time estimates from real environments. Seventeen tests in 12 environments were conducted using noise bursts. Decay rates were estimated using the MLE algorithm (ordinate) and the extrapolated T_{60} times were compared with estimates from Schroeder's method (abscissa). (a) Estimates of T_{60} in one-third octave bands with center frequencies exceeding 1 kHz (open circle) and their average (filled square). (b) Broad-band estimates of T_{60} from the recorded room response. Data shown by open circles are from different locations in an auditorium. Results are from one presentation of noise burst in each test.

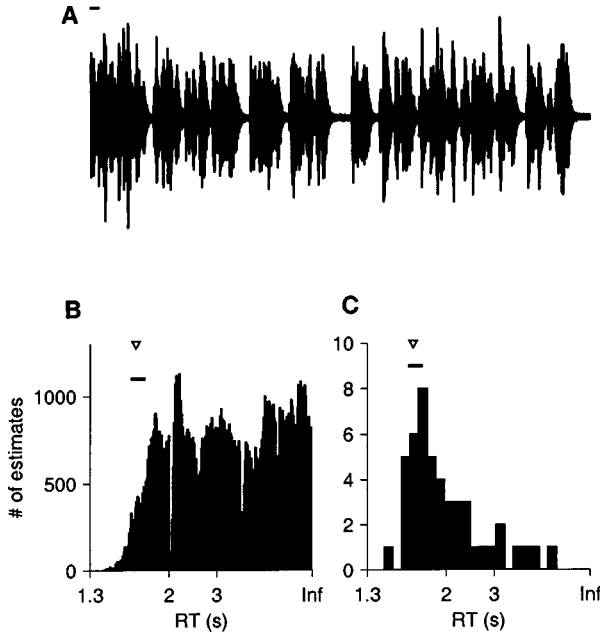

FIG. 11. Evaluation of room reverberation time (RT) from connected speech played back in a partially open circular foyer. The RT for this environment as measured from hand-claps was 1.66±0.07 s (Schroeder's method) and 1.62 s (from MLE procedure). (a) Trace of CST passage (duration 50 s) recorded in the environment. Bar indicates 1 s. (b) The histogram of MLE estimates over the duration of recording. The first peak in the aggregate histogram is the best RT estimate from connected speech (1.83 s). The horizontal bar is the range of RT estimates obtained from Schroeder's method, and the triangle indicates the MLE estimate. (c) Peak values from histogram of estimates were obtained every 1 s, and the 50 peak values were used to produce the histogram shown. The best estimate of RT from this histogram is at the dominant peak (1.7 s).

signals, except for one outlier. The outlier along with three other data points (open circles) were obtained in a large auditorium. The outlying data point was obtained at a source-to-microphone distance of 4 m, whereas the three remaining data points were obtained at a source-to-microphone distance of 1.5 m (at three different locations in the auditorium). The sound levels were not adjusted to compensate for the distance, and hence the experiment corresponding to the outlier was at a lower SPL, resulting in reduced dynamic range (from peak SPL to noise floor). For the four experiments in the auditorium, the Schroeder estimates of T_{60} (in seconds) were 2.18 (outlier), 0.39, 0.39, and 0.33, respectively. The MLE estimates, on the other hand, were 0.69 (outlier), 0.77, 0.80, and 0.67, respectively. Schroeder's method appears to be sensitive to the peak-SPL to noise-floor range, because the remaining three locations provided RT values that were in good agreement. On the other hand, the MLE estimates, while larger than the Schroeder estimates, were consistent and relatively robust irrespective of the source-to-microphone distance. That is, a reduction in dynamic range appears to affect estimates from the MLE to a lesser extent than estimates from Schroeder's method. A more detailed study is required to quantify the effect of dynamic range on the various estimation methods, and has not been attempted here.

These results raise the issue of estimation in narrow bands. It appears, although it is by no means conclusive, that the upper one-third octave bands (over 1 kHz) may provide more accurate estimates than the lower bands. Frequency decomposition is a standard part of most audio signal processing algorithms, and so it may be useful to track estimates in the higher frequency bands, or in select bands where the energy is greatest. Tracking high-energy bands is likely to provide more temporal range in tracking decays before encountering the noise floor, and thus sharpen the peak in the histogram of estimates. Alternatively, averaging over all high-frequency bands can provide estimates that are in closer agreement with T_{60} times obtained from Schroeder's method.

The above findings suggest that there is good correlation between the estimates obtained with the MLE procedure and those obtained with Schroeder's method. While Schroeder's method provides the most accurate estimates, in situations where the peak to noise-floor range is limited, the MLE method can provide robust estimation.

F. Estimation of RT from connected speech in real listening environments

The results presented in the preceding sections indicate that the MLE output is in good agreement with actual or simulated room RTs. In particular, the estimator can be ap-

plied to isolated word utterances, even though the naturally decaying offsets of terminating phonemes may lead to an overestimation of RT (see Fig. 6). Here, we test the performance of the procedure explicitly in a challenging estimation task, namely estimating room RT from connected speech.

A segment of speech (about 50 s in duration) from the Connected Speech Test (CST) corpus was played back in a partially open, circular foyer (one-third octave band analysis shown in Fig. 9, square symbols). The RT for this environment was first estimated with hand-claps using Schroeder's method (1.66 ± 0.07 s) and independently confirmed with the MLE procedure (estimated RT from histogram was 1.62 s). The MLE procedure was then applied to the recorded speech data [Fig. 11(a)]. A histogram of room decay rates for the duration of the recorded data was constructed [Fig. 11(b)]. The order-statistics filter was used to select the first dominant peak in the histogram (RT=1.83 s). This is the best RT estimate based on the aggregate data. It is possible to refine the procedure for arriving at the best estimate by applying the order-statistics filter at much shorter time intervals. Towards this end, a histogram was constructed at intervals of 1 s, and the best RT estimate for this interval was obtained. The resulting best estimates from all 1-s durations (50 in all) were binned to produce the histogram shown in Fig. 11(c). It can be seen that the number of estimates peaks at RT=1.7 s, which agrees with the mean value of 1.66 s from Schroeder's method (using hand-claps), and is well within its standard deviation [0.07 s; the one-sigma interval is indicated by the horizontal bar in Figs. 11(b) and (c)].

Given that terminal phonemes have a natural decay rate (see Fig. 6), it is not surprising that the MLE procedure produces estimates somewhat larger than the real room RT. Further, the discrepancy between the actual RT and those estimated from connected speech arise from the absence of adequate numbers, and the limited duration, of gaps [see Fig. 11(a)]. Thus, regions of free decay where estimation is accurate are limited. Not withstanding these constraints, the procedure works well, in part due to the decision-making capability built into the order-statistics filter. By selecting the first dominant peak (from the left) in the histogram, the filter in effect rejects spurious estimates, thereby reducing the error in the estimation procedure. The mean value of the histogram or its median, for instance, would result in significantly higher estimates of RT. The performance of the order-statistics filter can be further improved if one were to obtain a statistical characterization of gap duration from a large corpus of connected speech or other sounds. Such a characterization can provide a robust percentile cutoff value [see Eq. (12)] which could then be used to select the best RT value for the room (results not shown).

In conclusion, the MLE procedure, in combination with order-statistics filtering, provides a robust means for blind estimation of room RT. The procedure has been validated against Schroeder's method, and with real room data such as hand-claps, isolated word utterances, and connected speech.

V. DISCUSSION

The estimation of reverberation time is a widely investigated problem. Traditionally, two approaches have been taken. The RT is computed analytically using formulas that incorporate the geometry and absorptive characteristics of the reflecting surfaces, or empirically using a test sound with known properties that is radiated into the environment, and for which the RT is estimated from the received sounds. The former approach is embodied in the Sabine-type formulas (Sabine, 1922; Eyring, 1930; Millington, 1932; Sette, 1933; see Young, 1959; Kuttruff, 1991, for reviews), while the latter is based on the analysis of decay curves, such as using Schroeder's method (Schroeder, 1965a; Chu, 1978; Xiang, 1995). These methods have wide applicability and have been used extensively since they were developed. The current work complements these methods, and provides a technique for evaluating RT from passively received microphone signals.

In the Interrupted Noise Method the excitation signal is broadband or narrow-band filtered noise that is abruptly switched off at a known time, and is followed by a sufficiently long pause to track the decay. The reverberation experiment thus requires careful control of a known sound source (the excitation signal), and is repeated many times to arrive at an average RT estimate. In contrast, Schroeder's method eliminates multiple trials, and can be carried out with an impulsive sound, such as a pistol shot, to obtain a reliable RT estimate on a single trial. For narrow-band estimates, a filtered impulse can be used (see also footnote 1). For the interrupted noise method, the RT is determined by selecting a decay range, beginning 5 dB below the initial level at sound offset and going down a further 20 or 30 dB, taking care to remain above the noise floor. For this method, sound offset time should be known. In Schroeder's method, precise knowledge of the impulse occurrence time is not necessary, except that the decay range should be above the noise floor [see ISO 3382 (1997) for a discussion of this point]. When the impulsive sound is well-isolated, i.e., preceded and followed by a sufficiently long period of silence, and the background noise level is well outside the measured decay range, Schroeder's method will provide the best estimates of RT.

The motivation for developing the MLE procedure was to extend the utility of the decay curve method to situations where excitation signals are not available to conduct a reverberation experiment. Instead one has to rely on passively received microphone signals consisting of unknown sound segments with randomly occurring gaps. In such a blind situation, it is expected that the method will be less reliable than Schroeder's method, and so the goal was to combine a theoretically proven procedure (the maximum-likelihood approach) followed by a decision strategy that reduces the estimation error. It was hoped that such an approach would allow the estimator performance to approach that of Schroeder's method.

The MLE procedure is similar to the noise decay curve and Schroeder's methods. It differs from these methods in that it is parametric, and is based on a widely accepted model of the reverberant tail, namely the exponential decay model [see Young (1959) for a discussion on how the Sabine type formulas are related to a linear decay of the sound pressure level after the source is turned off]. The model assumes that

the amplitude of successive reflections are damped exponentially, while the fine structure is a random uncorrelated process. The random fine structure is a reasonable assumption because the excitation signal is random, and so the room output is also random. Schroeder makes this assumption explicit when developing his method (Schroeder, 1965a), arguing that the phase and amplitudes of the normal modes at the time of sound offset are unknown, and so the decaying normal modes (of different frequencies) constitute a random process even if the room response is deterministic. For most diffusive environments, this approximates the reverberant tail fairly well (see Fig. 1) and forms the central assumption of the work reported here.

The success of the MLE approach derives largely from the analytically tractable nature of the maximum-likelihood formulation, reducing the problem to the estimation of a single parameter that can be determined computationally. We also showed that for ongoing and onset segments of the sound, the estimates will assume implausible values as the model is not valid in these regions. However, an order-statistics filter downstream to the maximum-likelihood estimator can reject these estimates and extract the room RT with improved confidence. This is based on the intuitive idea that sounds cannot decay faster than the rate prescribed by the room decay rate, and thus selecting the earliest peak improves the confidence of the estimates. To our best knowledge, this MLE approach to blind decay rate estimation in enclosures has not been reported in the major acoustical literature.

The two encouraging results of this study are the validation of the estimates using Schroeder's method, and the RT estimates obtained from speech sounds. Under ideal conditions (impulsive hand-claps), the MLE method produced results that were comparable to Schroeder's method (Figs. 1 and 7), and provided motivation to carry out further tests using speech sounds. Speech sounds present particular problems to most estimation algorithms because they violate the two most commonly held assumptions, namely stationarity and Gaussian statistics. Further, even abruptly terminating phonemes such as stop consonants demonstrate a gradual decay, with a rate that may be in the range of 5–40 ms. Thus gradual offsets can increase the overall decay rate estimated in reverberant environments. However, except for the increase in estimated decay rate (a variation up to about 15% for sounds terminating in /d/), the tracking and histogram procedure works rather well, indicating that the method is relatively robust to model uncertainties.

Partially blind approaches to RT estimation have previously been described. (1) A neural network can be trained to learn the characteristics of room reverberation (Nannariello and Fricke, 1999; Cox *et al.*, 2001). Here, it is necessary to train the network whenever the environment changes. (2) The signal is explicitly segmented to identify gaps wherein decays can be tracked (Lebart *et al.*, 2001). It should be noted that the order-statistics filter developed in this work performs an implicit segmentation of the signal by rejecting estimates that are implausible. (3) A blind dereverberation procedure can be used to obtain the room impulse response. However, the room impulse response must be minimum phase, a condition that most listening environments fail to satisfy (Neely and Allen, 1979; Miyoshi and Kaneda, 1988).

The MLE procedure presented here is just one method for estimating room RT. Other methods are also possible. For instance the envelope of the sound can be extracted in the estimation interval, converted to sound pressure level, and a regression line could be fitted to obtain the T_{60} time. This is a blind version of the RT estimation procedure followed by Lebart *et al.* (2001). The order-statistics filter can be applied to the histogram of estimates as with the MLE procedure. The method is nonparametric and so is not subject to model uncertainties. This approach was used to estimate the decay rate of isolated word utterances (Fig. 6). While a detailed comparison of the methods is beyond the scope of this work, we note that the MLE procedure is a theoretically principled way of extracting the decay rate from the sound envelope.

The MLE procedure is model-based and is expected to perform reasonably well in diffuse sound fields (i.e., uniform with respect to directional distribution) and where a single decay rate describes the reverberant tail. For most sound fields this is a reasonable approximation [see Kuttruff (1991) for a discussion on this point]. The estimates of T_{60} are in good agreement with Schroeder's method in most of the listening environments tested, including challenging situations where the source or recording microphone was close to a wall, or there was moderate background noise (see Fig. 8). While the MLE procedure produces best results when there are isolated impulsive sounds or abruptly terminating white noise bursts, the results of tests with isolated word utterances and connected speech are in good agreement with the actual T_{60}. Thus, the procedure is expected to work under most listening conditions.

A result that was particularly interesting was the apparent robustness of the MLE method to reduced dynamic range of sounds (i.e., situations where the peak to noise floor range of the decay curve was small). The MLE method provided consistent estimates even when the dynamic range of sound decay was reduced. This is illustrated in Fig. 10(b) which shows the effect of source-to-microphone distance on RT estimation. The four open circles were broadband estimates obtained in an auditorium, of which one experiment (corresponding to the outlier) was at a larger source-to-microphone distance (4 m) than the remaining three (1.5 m) (see Sec. III for details). Broadly speaking, at constant sound level output from the source, the MLE method provided comparable estimates of RT including when the source-to-microphone distance ranged from 1.5 to 4 m, with reduced dynamic range of the sound decay curve. In contrast, Schroeder's method is dependent on the peak to noise floor range of the decay curve, and reducing the range can result in overestimation of the RT.[3] Consequently, increasing the source-to-microphone distance affected estimates for Schroeder's method more than those for the MLE method. The ISO recommendations for measurement using Schroeder's method specify that ''the level of the noise floor shall be at least 10 dB below the lower value of the evaluation range'' (ISO 3382, 1997). For example, if a 20-dB range is to be used, then the recommended peak to noise floor range must be at least 35 dB (including the initial 5-dB response from peak). This finding

should be interpreted with caution because the MLE method was not tested in high levels of background noise or when the dynamic range was drastically reduced. The method appears to be more robust than Schroeder's method only for the conditions tested here. To properly evaluate this effect two lines of inquiry need to be pursued: (1) quantify the effect of source-to-microphone distance on the RT estimates from the two methods, and (2) explicitly incorporate additive background noise in the MLE procedure. Incorporating background noise would require the estimation of an additional parameter, the power of the background noise. This would help to determine more precisely the relative merits of the different methods, and, in particular, to indentify situations where the MLE method can provide improvements over Schroeder's method.

The method proposed here can be expected to perform poorly when there are room resonances and the sound pressure level decays nonlinearly with time. This can be a result of the room geometry, or positioning the recording microphone in a region of the sound field that is nondiffusive, or in acoustically coupled spaces with widely differing RTs. In addition to model failure, the performance of the estimator may be poor when there are insufficient numbers of gaps, or there is fluctuating background noise. Good performance results when there are about 10% gaps and the peak sound level (at the time of offset) is about 25 dB SPL over the noise floor. Performance may also be compromised when background noise is modulated (such as with background music or babble) as the procedure will attempt to track any modulation present in the environment, and hence produce multimodal histograms with peaks that may not be easily discriminated.

The blind estimation procedure suggested here can be applied in a number of situations. Because only passive sounds are used, any audio processor that has access to microphone input can estimate the room reverberation time, either in single-channel (broadband) or multi-channel (narrow-band) mode. Further, while the method presented here is for a single microphone, it can be applied with no modifications to an array of microphones, providing several independent estimates of the RT. One of the most interesting applications is in the selection of signal processing strategies tailored to specific listening environments. These include hearing aids and hands-free telephony. Programmable hearing aids often have the ability to switch between several processing schemes depending on the listening environment (Allegro *et al.*, 2001). For instance, in highly diffusive environments, where the source-to-listener distance exceeds the critical distance, adaptive beamformers are ineffective (Greenberg and Zurek, 2001). In such situations, it would be convenient to switch off the adaptive algorithm and revert to the relatively simple (fixed) delay-and-sum beamformer. Alternatively, in highly confined listening environments such as automobile interiors, where a reflecting surface is located in close proximity to the ear, it may be convenient to switch-off the proximal ear microphone, and use the input from the microphone located in the better (more distal) ear. Such decisions can be made if there is a passive method for determining reverberation characteristics. Other potential applications could include hands-free telephony, and room acoustics evaluation in sound-level meters. A limitation of the method is its relatively poor performance with narrow-band signals whose center frequencies are below 1 kHz. However, the performance is good for broadband signals, and narrow-band signals whose center frequencies exceed 1 kHz.

The computational costs of implementing the procedure are largely due to the iterative solution of the maximum-likelihood equation. We have developed fast algorithms for reducing the computational cost so that the procedure can be implemented in real-time (Ratnam *et al.*, 2003). Thus, the method can be implemented in passive listening devices to determine the reverberation characteristics of the environment.

ACKNOWLEDGMENTS

We would like to thank the members of the Intelligent Hearing Aid Project at the Beckman Institute, University of Illinois, for their criticisms and comments at various stages of the work. The constructive comments and suggestions of the anonymous referee helped greatly in improving the clarity of the manuscript, and in providing the necessary perspective for evaluating the work with respect to Schroeder's method. The work was supported by grants from the National Institutes of Health (R21DC04840), Phonak AG, Charles M. Goodenberger Foundation, and the Beckman Institute.

APPENDIX: CRÁMER–RAO BOUNDS FOR DECAY RATE ESTIMATION

Bounds on the estimate of a and σ are obtained from the variance of the score function, also called the Fisher information J. This is more conveniently expressed in terms of the derivatives of the score functions (Poor, 1994). Given the parameter $\boldsymbol{\theta}^{\mathrm{T}}=[a\,\sigma]$ and the score function $s_{\theta}^{\mathrm{T}}(\mathbf{y};\boldsymbol{\theta})=[s_a(\mathbf{y};a,\sigma)\,s_{\sigma}(\mathbf{y};a,\sigma)]$, we have

$$J(\boldsymbol{\theta})=-\mathbf{E}\left[\frac{\partial s_{\boldsymbol{\theta}}^{\mathrm{T}}(y;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]. \tag{A1}$$

From Eqs. (7), (9), and (A1), we have

$$J(\boldsymbol{\theta})=\begin{pmatrix}\dfrac{N(N-1)(2N-1)}{3a^2} & \dfrac{N(N-1)}{a\sigma}\\[2ex] \dfrac{N(N-1)}{a\sigma} & \dfrac{2N}{\sigma^2}\end{pmatrix}. \tag{A2}$$

By the Crámer–Rao theorem (Poor, 1994), a lower bound on the variance of any unbiased estimator is simply $J^{-1}(\boldsymbol{\theta})$, which is

$$J^{-1}(\boldsymbol{\theta})=\begin{pmatrix}\dfrac{6a^2}{N(N^2-1)} & -\dfrac{3a\sigma}{N(N+1)}\\[2ex] -\dfrac{3a\sigma}{N(N+1)} & \dfrac{\sigma^2(2N-1)}{N(N+1)}\end{pmatrix}. \tag{A3}$$

From the asymptotic properties of maximum-likelihood estimators (Poor, 1994), we know that the estimates of a and

σ are asymptotically unbiased and their variances achieve the Crámer–Rao lower bound (i.e., they are efficient estimates). Thus, if a^* and σ^* are the estimates obtained from the solutions of Eqs. (8) and (11), the variance of the estimates are

$$\mathbf{E}[(a^*-a)^2] \geqslant \frac{6a^2}{N(N^2-1)}, \tag{A4}$$

$$\mathbf{E}[(\sigma^*-\sigma)^2] \geqslant \frac{\sigma^2(2N-1)}{N(N+1)}, \tag{A5}$$

with equality being achieved in the limit of large N. As the variance of a and σ are $O(N^{-3})$ and $O(N^{-1})$, the estimation error can be made arbitrarily small if observation windows are made sufficiently large.

[1]An interesting discussion on the use of excitation signals can be found in an exchange of letters between Smith (1965) and Schroeder (1965b). Schroeder uses the term ''tone-burst'' to denote a filtered impulse having a narrow-band, and this appears to have caused some confusion.

[2]The offset time courses for speech that are reported here are the results of analyzing isolated word utterances. These are the authors' unpublished observations based on a preliminary study. The offset time courses of speech segments in connected speech and connected discourse require further study.

[3]The noise floor of the integrated impulse response curve (Schroeder's method) manifests itself as a relatively flat line. When the decay range for measuring RT includes a portion of the noise floor, the estimated RT will be greater than the true value [see ISO 3382 (1997) for a discussion on this point].

Allegro, S., Buechler, M., and Launer, S. (**2001**). ''Automatic sound classification inspired by auditory scene analysis,'' Proc. European Conf. Sig. Proc., EURASIP.

Bolt, R. K., and MacDonald, A. D. (**1949**). ''Theory of speech masking by reverberation,'' J. Acoust. Soc. Am. **21**, 577–580.

Chu, W. T. (**1978**). ''Comparison of reverberation measurements using Schroeder's impulse method and decay curve averaging method,'' J. Acoust. Soc. Am. **63**, 1444–1450.

Cox, R. M., Alexander, G. C., and Gilmore, C. (**1987**). ''Development of the connected speech test (CST),'' Ear Hear. **8**, 119–126.

Cox, T. J., Li, F., and Darlington, P. (**2001**). ''Extracting room reverberation time from speech using artificial neural networks,'' J. Audio Eng. Soc. **49**, 219–230.

Eyring, C. F. (**1930**). ''Reverberation time in 'dead' rooms,'' J. Acoust. Soc. Am. **1**, 217–241.

Greenberg, J. E., and Zurek, P. M. (**2001**). ''Microphone-array hearing aids,'' in *Microphone Arrays: Signal Processing Techniques and Applications*, edited by M. Brandstein and D. Ward (Springer-Verlag, Berlin), pp. 229–253.

Hartmann, W. M. (**1997**). ''Listening in a room and the precedence effect,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, New York), pp. 191–210.

International Phonetic Association (**1999**). *Handbook of the International Phonetic Association* (Cambridge U.P., Cambridge). The American-English sound files are available on-line at ⟨http://uk.cambridge.org/linguistics/resources/ipahandbook/american-English.zip⟩.

ISO 3382 (**1997**). *Acoustics—Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters*, 2nd ed. (International Organization for Standardization, Géneve).

Knudsen, V. O. (**1929**). ''The hearing of speech in auditoriums,'' J. Acoust. Soc. Am. **1**, 56–82.

Kuttruff, H. (**1991**). *Room Acoustics*, 3rd ed. (Elsevier Science Publishers Ltd., Lindin).

Lebart, K., Boucher, J. M., and Denbigh, P. N. (**2001**). ''A new method based on spectral subtraction for speech dereverberation,'' Acustica **87**, 359–366.

Millington, G. (**1932**). ''A modified formula for reverberation,'' J. Acoust. Soc. Am. **4**, 69–82.

Miyoshi, M., and Kaneda, Y. (**1988**). ''Inverse filtering of room impulse response,'' IEEE Trans. Acoust., Speech, Signal Process. **36**, 145–152.

Nabalek, A. K., and Pickett, J. M. (**1974**). ''Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners,'' J. Speech Hear. Res. **17**, 724–739.

Nabalek, A. K., and Robinson, P. K. (**1982**). ''Monaural and binaural speech perception in reverberation for listeners of various ages,'' J. Acoust. Soc. Am. **71**, 1242–1248.

Nabalek, A. K., Letowski, T. R., and Tucker, F. M. (**1989**). ''Reverberant overlap- and self-masking in consonant identification,'' J. Acoust. Soc. Am. **86**, 1259–1265.

Nannariello, J., and Fricke, F. (**1999**). ''The prediction of reverberation time using neural network analysis,'' Appl. Acoust. **58**, 305–325.

Neely, S. T., and Allen, J. B. (**1979**). ''Invertibility of room impulse response,'' J. Acoust. Soc. Am. **66**, 165–169.

Pitas, I., and Venetsanopoulos, A. N. (**1992**). ''Order statistics in digital image processing,'' Proc. IEEE **80**, 1893–1921.

Poor, V. (**1994**). *An Introduction to Signal Detection and Estimation* (Springer-Verlag, New York).

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (**1992**). *Numerical Recipes in C* (Cambridge U.P., Cambridge).

Ratnam, R., Jones, D. L., and O'Brien, Jr., W. D. (**2003**). ''Fast algorithms for the blind estimation of reverberation time,'' IEEE Signal Process Lett. (in press).

Sabine, W. C. (**1922**). *Collected Papers on Acoustics* (Harvard U.P., Cambridge).

Schroeder, M. R. (**1965a**). ''New method for measuring reverberation time,'' J. Acoust. Soc. Am. **37**, 409–412.

Schroeder, M. R. (**1965b**). ''Response to 'Comments on New method of measuring reverberation time,' '' J. Acoust. Soc. Am. **38**, 359–361(L).

Schroeder, M. R. (**1966**). ''Complementarity of sound buildup and decay,'' J. Acoust. Soc. Am. **40**, 549–551.

Sette, W. J. (**1933**). ''A new reverberation time formula,'' J. Acoust. Soc. Am. **4**, 193–210.

Smith, Jr., P. W. (**1965**). ''Comment on 'New method of measuring reverberation time,''' J. Acoust. Soc. Am. **38**, 359(L).

Tahara, Y., and Miyajima, T. (**1998**). ''A new approach to optimum reverberation time characteristics,'' Appl. Acoust. **54**, 113–129.

Xiang, N. (**1995**). ''Evaluation of reverberation times using a nonlinear regression approach,'' J. Acoust. Soc. Am. **98**, 2112–2121.

Young, R. W. (**1959**). ''Sabine reverberation equation and sound power calculations,'' J. Acoust. Soc. Am. **31**, 912–921.