Reliability of estimating the room volume from a single room impulse response

Martin Kuster[a)]

Laboratory of Acoustic Imaging and Sound Control, Delft University of Technology, 2600 GA Delft, The Netherlands

(Received 23 January 2008; revised 13 May 2008; accepted 13 May 2008)

The methods investigated for the room volume estimation are based on geometrical acoustics, eigenmode, and diffuse field models and no data other than the room impulse response are available. The measurements include several receiver positions in a total of 12 rooms of vastly different sizes and acoustic characteristics. The limitations in identifying the pivotal specular reflections of the geometrical acoustics model in measured room impulse responses are examined both theoretically and experimentally. The eigenmode method uses the theoretical expression for the Schroeder frequency and the difficulty of accurately estimating this frequency from the varying statistics of the room transfer function is highlighted. Reliable results are only obtained with the diffuse field model and a part of the observed variance in the experimental results is explained by theoretical expressions for the standard deviation of the reverberant sound pressure and the reverberation time. The limitations due to source and receiver directivity are discussed and a simple volume estimation method based on an approximate relationship with the reverberation time is also presented.
 [DOI: 10.1121/1.2940585]

PACS number(s): 43.55.Gx, 43.55.Br, 43.55.Mc [NX] Pages: 982–993

I. INTRODUCTION

The general course in room acoustics research is to compare measurements of the room impulse response (RIR) or total sound pressure level with predictions obtained from room acoustic models using geometrical and acoustical room parameters.[1,2] An interesting problem is to reverse the process and observe to what extent and accuracy these parameters can be retrieved from measured data. Depending on the approach followed, this fits into the subjects of inverse methods or parameter extraction. In a room acoustics context, the most important geometry parameters are the room volume and the source-to-receiver distance. In the present paper the focus is on the estimation of the former, but the estimation of the latter is also investigated.

The ease and accuracy with which the room volume can be estimated from a single RIR is relevant to the understanding of room acoustics for the following reasons. The (combination of) parameters extracted from the RIR for the volume estimation are those that do change with a change in room volume. In this context it is interesting to note that a number of perceptual experiments performed by Cabrera and colleagues indicated that auditory room size perception is related to clarity index.[3,4] Further, if it proves to be very difficult to obtain accurate volume estimates, it can be concluded that the exact value of this parameter does not greatly affect the RIR. Apart from the relevance to basic room acoustics research, the estimation of the room volume by acoustic means can have practical applications in cases where, for a number of possible reasons, the room volume cannot be determined by other means.

At least three possible approaches can be identified for the estimation of the room parameters. The first approach requires geometric arrays of receiver positions and was shown previously to provide detailed room information but cannot be used with a single RIR.[5] The second approach is based on the extraction of acoustic parameters from a RIR that are then used inversely with one of the standard room acoustic models. A number of authors have used this approach to find a, not necessarily unique, optimum room parameter set that, when fed into the room acoustic model, results in the desired target values for the acoustic parameters.[6,7] The third approach is based on the extraction of more general signal parameters, of which the acoustic parameters may be a subset, that are then used in conjunction with "blind" methods such as maximum-likelihood or neural networks. The extraction of suitable signal parameters for the parametrization of RIRs has been performed by Hulsebos[8] and van der Vorm,[9] but the found parameters are not applicable to the estimation of geometrical room parameters. Blind methods in room acoustics have been investigated for example by Li and Cox[10] and Ratnam *et al.*[11] In the present paper, the second approach is followed because it can be used with a single RIR and has the potential of using the limited available data more effectively than the third approach. The three room acoustic models employed are based on geometrical acoustics, eigenmode or diffuse field assumptions. The suitability of each model and consequent success of the estimation method is considered separately.

The framework, within which the estimation methods are to be applied, is as follows. No knowledge is to be assumed about either the source or receiver characteristics, their position within the room, or the distance between them. Further, no assumptions are employed about the acoustic characteristics of the room. Several limitations that are

[a)]Electronic mail: kuster_martin@hotmail.com.

 0001-4966/2008/124(2)/982/12/$23.00

TABLE I. List of measured rooms together with the number of receiver positions N_R, the geometric room volume V_{geo}, the broadband reverberation time T_{60}, the absorption area A, and the Schroeder frequency $f_{Schroeder}$.[a]

Name (location)[a]	N_R	V_{geo}[b] (m^3)	T_{60} (s)	A (m^2)	$f_{Schroeder}$ (Hz)	Shape, Remarks
Lavatory (SARC, QUB)	18	5	0.3	2.6	490	Rectangular, few absorption
Office (LG023, SARC, QUB)	15	60	0.6	16	200	Rectangular, corner protrusion
Listening room (LG013, SARC, QUB)	10	131	0.3	70	96	Rectangular, special treatment
Multimedia room (SARC, QUB)	14	150	0.4	60	103	Rectangular, corner protrusion
Lecture hall A (Zaal G, TU Delft)	143	180	0.9	32.2	141	Rectangular, tiered seating
Lecture hall B (School of Music, QUB)	4	550*	1.0	89	85	Rectangular with bay window
McMordie Hall (School of Music, QUB)	9	850*	1.4	98	85	Rectangular plan, roofed ceiling
Harty Room (School of Music, QUB)	18	1 150*	1.4	132	70	Stage, side choirs, roofed ceiling
Sonic Laboratory (SARC, QUB)	30	3 200	0.7	736	30	Rectangular, grid floor at 4 m
Whittla Hall (QUB)	8	8 400*	1.8	751	30	Rectangular plus stage house
Concert hall A (Concertgebouw Amsterdam)	420	19 000	2.6	1180	23	Horseshoe, columns, balconies
Concert hall B (De Doelen, Rotterdam)	512	24 000	2.3	1760	19	Irregular, inner shell

[a]TU=Delt University of Technology, QUB=Queen's University Belfast, and SARC=Sonic Arts Research Centre.
[b]An asterisk indicates that it has been determined from imcomplete dimension data and may be inaccurate.

caused by these conditions or the possibilities that arise from a relaxation thereof are discussed at various stages in the present paper.

II. DESCRIPTION OF ROOMS INVESTIGATED

A summary of all rooms included in the investigation is given in Table I. In Table I, V_{geo} is the value of the room volume obtained from architectural drawings and/or geometrical measurements. In the rooms with a suspended ceiling, the first figure is the measurement up to the acoustic ceiling and the figure in parentheses is the approximate volume up to the fixed ceiling.

In lecture hall A, the room impulse responses have been measured using the maximum-length sequence method with a polyhedral loudspeaker (designed to produce omnidirectional radiation over a wide frequency bandwidth) and a sampling frequency of 14 980 Hz. The measurements in concert halls A and B have been performed with the same polyhedral loudspeaker but using a logarithmic sweep and a sampling frequency of 16 kHz. In the remaining rooms, the measurements were performed using a logarithmic sweep with a Mackie HR 824 loudspeaker (a commercial studio monitor) as a sound source and a sampling frequency of 48 kHz. The measurement microphone for all rooms was the omnidirectional channel of a SoundField MKV microphone system.

In lecture hall A and concert halls A and B, the measurement positions described a line across the entire width of the hall and the offset parameter used in some of the figures is the distance from the central receiver position. In the remaining rooms a varying number of representative receiver positions have been selected. Only one source position was used in all rooms.

III. GEOMETRICAL ACOUSTICS METHOD

Within geometrical acoustics, sound waves are replaced by sound rays and reflected waves are replaced by (specular) reflections. The temporal density of reflections is given by[12]

$$\frac{dN_t}{dt} = 4\pi\frac{c^3t^2}{V}, \tag{1}$$

and therefore depends on time t, the room volume V, and the wave speed c in air. The wave speed does not vary considerably within a practical temperature range and can be considered to be known. Equation (1) can thus be rearranged to yield the volume from the temporal reflection density.

In a modeled RIR $h(t)$ the reflections can be identified because it is assumed here that they are represented by scaled Kronecker delta functions. The binary signal $h_{refl}(t)$ is then constructed from $h(t)$ in the following manner:

$$h_{refl}(t) = \begin{cases} 0, & \forall |h(t)| = 0 \\ 1, & \forall |h(t)| \neq 0. \end{cases} \tag{2}$$

As illustrated in Fig. 1(b) for the example RIR in Fig. 1(a), $h_{refl}(t)$ is essentially a pulse width modulated signal with the modulation density equal to the reflection density. An esti-

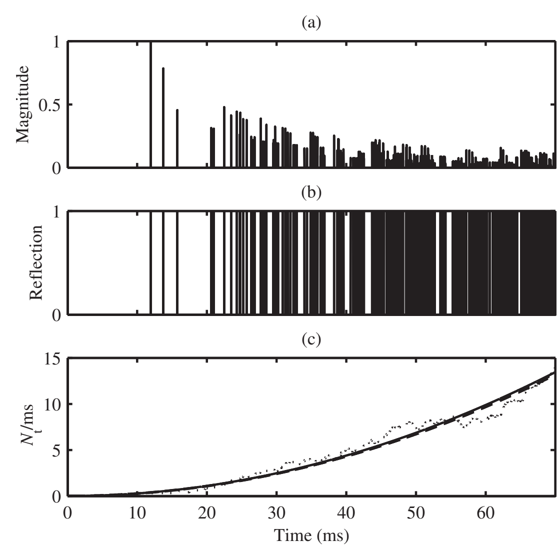

FIG. 1. (a) Modeled RIR $h(t)$, (b) identified reflections in $h_{refl}(t)$, and (c) estimated (dotted curve) and theoretical (solid curve) reflection density dN_t/dt. The almost coincident dashed curve is the least-squares fit to the dotted curve.

mate of the latter can be obtained after convolving $h_{\mathrm{refl}}(t)$ with a low pass moving average filter $f(t)$,

$$dN_t/dt \approx h_{\mathrm{refl}}(t) * f(t), \tag{3}$$

with the $(*)$ the convolution operator and the filter $f(t)$ of length T given by

$$f(t) = \begin{cases} 1/T, & \forall -T/2 \leqslant t \leqslant T/2 \\ 0, & \forall -T/2 > t > T/2. \end{cases} \tag{4}$$

The dotted curve in Fig. 1(c) shows the estimated reflection density, which clearly exhibits deviations from the theoretical behavior indicated by the solid curve. Because it is known that the density increases with the square of time, a least-squares fit can be applied to the estimated density and the result is the dashed curve in Fig. 1. For this particular example, there is virtually no difference between this curve and the theoretical reflection density given by the solid curve.

In order to assess the performance of this proposed volume estimation method, RIRs have been modeled with a mirror image source model (discrete time, frequency-independent reflection coefficient[13]) in rectangular rooms ranging in volume V from 10 to 10 000 m^3. The room dimensions (L_x, L_y, L_z) are given by

$$L_x = \varphi_x \left(\frac{V}{\varphi_x \varphi_y} \right)^{1/3}, \quad L_y = \varphi_y \left(\frac{V}{\varphi_x \varphi_y} \right)^{1/3}, \tag{5a}$$

$$L_z = \left(\frac{V}{\varphi_x \varphi_y} \right)^{1/3}, \tag{5b}$$

with φ_x a random variable with uniform distribution between 1.2 and 2 and φ_y a random variable with uniform distribution between 0.5 and 0.83. This procedure ensures that the aspect ratio of the room dimensions varies between 1.2:1:0.83 and 2:1:0.5. A similar procedure has been used for the positioning of the source and receiver within the room. The sampling frequency was 192 kHz and the reflections coefficient of all six walls was set to 0.6.

Due to the large range in room volumes, estimating the reflection density within a fixed time interval is prone to errors because small rooms have a very large density at the upper time limit and large rooms have a very small density at the lower time limit. The reflection density was thus estimated over a varying time interval defined by the arrival times of the first 500 reflections. The length T of the moving average filter was set to 20 ms. Informal experiments have shown that a variation by a factor of 2 on either side is acceptable.

Using these parameters, the room volume determined from the reflection density has been estimated for a total of 2000 modeled rooms. The mean error between true and estimated room volume was found to be 3.8% with a standard deviation of 5.7%. This numerical result shows that the room volume can be estimated fairly accurately under idealized conditions. The success of the method depends crucially on the reliable estimation of all individual reflections; this issue is now investigated further both theoretically and experimentally in the following two sections.

A. Resolution limit in the time domain

One necessary but not sufficient condition for the reflections to be represented by Kronnecker deltas is that the source impulse has infinite frequency bandwidth. In practice, this condition can never be met. Instead, a source impulse is now considered whose frequency response is of uniform magnitude and zero phase up to a maximum frequency $\omega_{\max}$. Its transfer function $H_{\mathrm{Sc}}(\omega)$ can thus be written as

$$H_{\mathrm{Sc}}(\omega) = \begin{cases} 1 & \text{for} |\omega| \leqslant \omega_{\max} \\ 0 & \text{for} |\omega| > \omega_{\max}. \end{cases} \tag{6}$$

From standard Fourier theory, its impulse response $h_{\mathrm{Sc}}(t)$ follows as

$$h_{\mathrm{Sc}}(t) = \frac{1}{\pi} \frac{\sin \omega_{\max} t}{t}. \tag{7}$$

The bandlimited RIR is then obtained by convolving $h(t)$, containing the scaled Kronnecker deltas, with $h_{\mathrm{Sc}}(t)$.

A condition is now required that specifies when two reflections arriving successively in time are separable and thus identifiable in the room impulse response. For this purpose, the Rayleigh resolution criterion is adopted.[14] It states that two impulses are barely resolved if the maximum of the first is located at the first zero of the second impulse. With $h_{\mathrm{Sc}}(t)$, this occurs when $\omega_{\max}\Delta t = \pi$ or

$$\Delta t = 1/2 f_{\max}, \tag{8}$$

where $f_{\max} = \omega_{\max}/2\pi$. An additional requirement is that the two impulses are $\pi/2$ out of phase. In the context of room acoustics, the phase differences are mainly caused by the imaginary part of the reflection coefficient of the walls and the directionally varying impulse response of source and receiver. Depending on these factors, the resolution criterion is either an over- or underestimate.

The inverse of the reflection density in Eq. (1) is the (average) time interval Δt between the arrival of successive reflections. Equating it with Δt from Eq. (8) results in a maximum time t_{up} up to which individual reflections are distinguishable,

$$t_{\mathrm{up}} = \sqrt{\frac{V f_{\max}}{2\pi c^3}}. \tag{9}$$

This result shows that the larger the volume V and maximum frequency $f_{\max}$, the larger the value t_{up} can assume.

Unfortunately, Eq. (1) and therefore also Eq. (9) are not directly applicable to RIRs measured in real rooms because the expressions have been derived for specular reflections and neglect the effects of scattering from rough surfaces and diffraction from finite-size surfaces. In order to circumvent this, $f_{\max}$ must be decreased to a, room dependent, value whose corresponding acoustic wavelength is much larger than the surface roughness and the size of the reflecting surfaces.

As practical examples, lecture hall A and concert hall A are considered and $f_{\max}$ is determined from a visual inspection of the room itself or photographs thereof. For lecture hall A, it is assumed that the smallest acoustic wavelength should be 30 cm and thus $f_{\max} \approx 1100$ Hz, which implies fur-

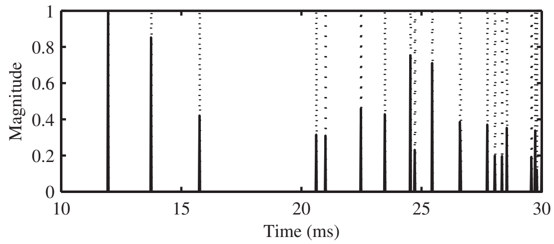

FIG. 2. Magnitude of modeled RIR $h(t)$ at the central receiver position in lecture hall A. The vertical dotted lines represent the signal $h_{\text{peaks}}(t)$.

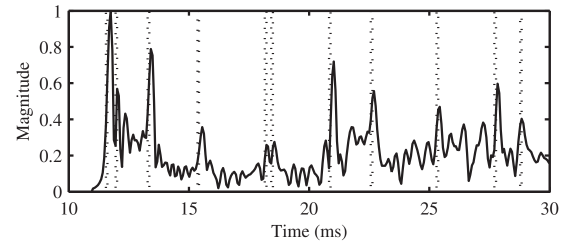

FIG. 3. Magnitude of measured RIR $h(t)$ at the central receiver position in lecture hall A. The vertical dotted lines represent the signal $h_{\text{peaks}}(t)$.

ther that $t_{\text{up}}=28$ ms and approximately 20 reflections should be distinguishable. For concert hall A, it is assumed that the smallest acoustic wavelength should be 2.0 m and thus $f_{\max}\approx 175$ Hz, which then implies that $t_{\text{up}}=110$ ms and approximately nine reflections should be distinguishable.

Note that the expression for t_{up} has been derived under the assumption that the real part of the reflection coefficient of all reflecting surfaces is approximately equal and that both source and receiver have omnidirectional directivity.

B. Identification of reflections in measured RIRs

In Eq. (2), the signal $h_{\text{refl}}(t)$ was formed by identifying the reflections as the only nonzero samples in the RIR $h(t)$. With measured RIRs this approach cannot be followed because the RIR magnitude is almost always nonzero. Instead, the extraction of the peaks in measured RIRs is now performed and it will then be considered whether the peaks correspond to the desired specular reflections. One way of extracting the peaks in a RIR is through adaptive thresholding known from image processing, see, e.g., Gonzales.[15] The rationale is that the magnitude of a peak is a factor ϵ above the magnitude average of a number of neighboring samples.

Using the mean as the method of averaging, the local magnitude mean at time t is given by

$$\mu_{\text{local}}(t)=\frac{1}{T_{\mu_{\text{local}}}}\int_{t-T_{\mu_{\text{local}}}/2}^{t+T_{\mu_{\text{local}}}/2}|h(\tau)|d\tau, \tag{10}$$

where $T_{\mu_{\text{local}}}$ is the averaging time. The binary signal containing the extracted peaks then follows as

$$h_{\text{peaks}}(t)=\begin{cases}0, & \forall h(t)<\epsilon\mu_{\text{local}}(t)\\ 1, & \forall h(t)\geqslant\epsilon\mu_{\text{local}}(t),\end{cases} \tag{11}$$

with ϵ the thresholding parameter. In applying this method to modeled and measured RIRs, the average was taken over $T_{\mu_{\text{local}}}=2$ ms and the threshold was set to $\epsilon=2$. Both values were determined experimentally and, as will be shown in the following, may not be optimal for all RIRs.

As a first step, the adaptive thresholding is applied to the RIR at the central receiver position in lecture hall A that was modeled with the mirror image source model. Together with the magnitude of the RIR, the result after thresholding is shown by the vertical dotted lines in Fig. 2 and it can be seen that all the present peaks corresponding to specular reflections are correctly identified. In this case, $h_{\text{peaks}}(t)=h_{\text{refl}}(t)$.

In Fig. 3, the result is shown if the same procedure is applied to the magnitude of the RIR measured at the same position in lecture hall A. Compared to Fig. 2, the number of extracted peaks is less. This is a consequence of the higher complexity of a measured RIR caused by the detailed impulse response of source and receiver, complex-valued, frequency-dependent reflection coefficients, the presence of nonspecular reflections, and other factors. On the other hand, the two identified peaks between 18 and 20 ms are not present in the modeled RIR. They are most likely caused by reflections from objects not modeled with the mirror image source model.

For a more comprehensive performance assessment, the peak extraction procedure has been performed on the RIRs measured across the width of lecture hall A and compared to the modeled RIRs at the same receiver positions. The two sets of RIRs are shown in Figs. 4(a) and 4(b), respectively. The result after applying adaptive thresholding to the magnitude of the measured RIRs is shown in Fig. 4(c). Compared to Fig. 4(a), it can be observed that the main features are reproduced. The deviations are again caused by differences between measured and modeled RIRs and in particular because some of the reflections in the measured RIR are weaker and more diffuse (mainly from the right sidewall)

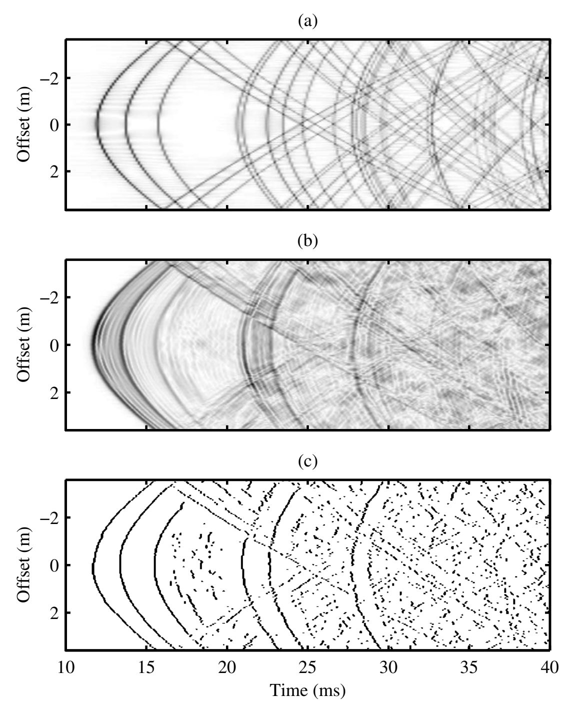

FIG. 4. RIR across the width of lecture hall A, (a) from model, (b) magnitude from measurements, and (c) result after applying adaptive thresholding to (b).

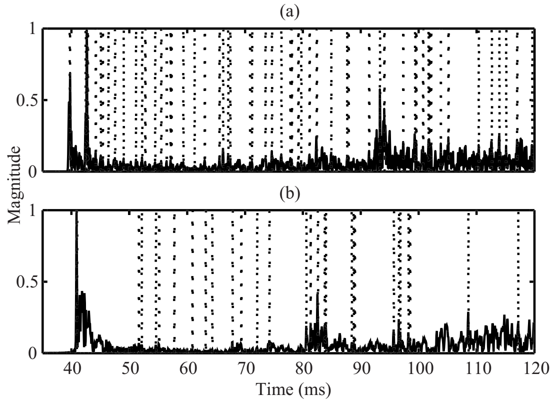

FIG. 5. Measured RIR in (a) concert hall A and (b) concert hall B with vertical dotted lines representing the signal $h_{\text{peaks}}(t)$.

and the measured RIR contains reflections from objects that are not present in the modeled RIR (mainly seen between 17 and 22 ms). Depending on the time interval in question, it seems fair to conclude that approximately 50% of the extracted peaks correlate with the specular reflections in Fig. 4(a). It has been contemplated in Sec. III A that for this room 20 reflections should be identifiable for a time interval up to 28 ms and the agreement is indeed best in this time interval. But the number of extracted peaks is less than 10.

Another issue is whether the adaptive thresholding parameters $T_{\mu_{\text{local}}}$ and ϵ found suitable for the RIRs in lecture hall A will yield usable results for other rooms as well. For this purpose, the same procedure has been applied to a measured RIR in both concert halls A and B and the results can be seen in Fig. 5. In both graphs the number of identified peaks is very large and most of them do not seem to correspond to specular reflections. From geometrical considerations, for concert hall A two sidewall reflections should arrive at 95 and 100 ms and for concert hall B one strong reflection should arrive at 85 ms and approximately another five reflections between 100 and 120 ms. The number of extracted peaks is therefore far too high. This observation is also supported by the theoretical result from Sec. III A that predicted for concert hall A nine identifiable reflections in the time interval up to 110 ms.

From these experimental results it is now concluded that systematically identifying specular reflections from measured RIRs is not a feasible approach in most rooms encountered in practice. Apart from the direct sound and depending on the acoustic characteristics of the room, at most one to five reflections can be extracted with confidence. As the room volume estimation through the reflection density explicitly relies on identifying the reflections, it is further concluded that this method is not a viable approach. It must also be noted that the theoretical expression for the reflection density is inaccurate for values of the time variable where the first few reflections arrive.[12]

IV. EIGENMODE METHOD

Whereas the approach for the room volume estimation presented in the previous section was based on the RIR in the time domain, the current section focuses on the room transfer function (RTF) in the frequency domain. Analogous to the reflection density, a possible approach for the room volume estimation is through the modal density. This requires the identification of individual eigenmodes and, since only the density of oblique modes is proportional to the room volume,[16] will need to be performed in a frequency region where rooms with at least a moderate amount of absorption do normally feature significant frequency overlap between neighboring eigenmodes (i.e., above the Schroeder frequency). For this reason, this approach will not be followed further in the current paper. Instead, it is noted that the Schroeder frequency is defined as[17,18]

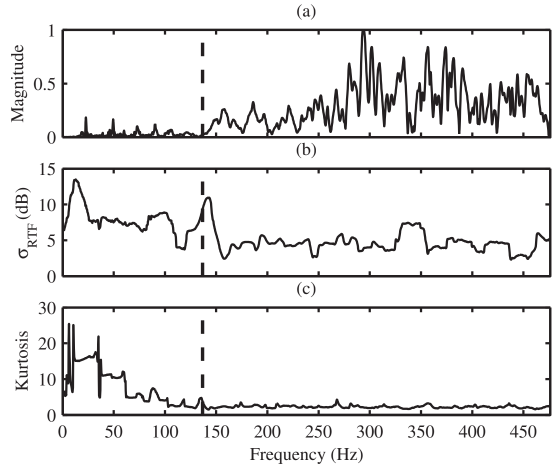

FIG. 6. (a) RTF magnitude in lecture hall A, (b) standard deviation of the logarithmic magnitude, and (c) kurtosis of the magnitude. The vertical dashed line indicates the theoretical Schroeder frequency.

$$f_{\text{Schroeder}} \approx 2000\sqrt{\frac{T_{60}}{V}} \tag{12}$$

and depends only on the room volume and the reverberation time T_{60}. The reverberation time can be estimated reliably and if it proves possible to estimate the Schroeder frequency experimentally from a RTF, the theoretical expression for the Schroeder frequency can be rearranged to yield an estimate of the room volume.

A. RTF statistics around the Schroeder frequency

Above the Schroeder frequency, the RTF magnitude statistics has a kurtosis of three (theoretical value for a Gaussian distribution) and a standard deviation of 5.57 dB.[19,17] It is anticipated that these two statistical parameters assume different values in the region below the Schroeder frequency and in order to verify this, they have been estimated on the RTFs in a sliding rectangular frequency window of 24 Hz width for a frequency range from 0 to 500 Hz. The measured RIRs were either zero-padded or truncated before transformation to the frequency domain such that the frequency resolution was always 0.3 Hz and therefore the statistics are estimated over 80 samples. The direct sound component was always included.

In Fig. 6, the results for lecture hall A show that both the standard deviation and the kurtosis fluctuate around the known statistical values above the Schroeder frequency,

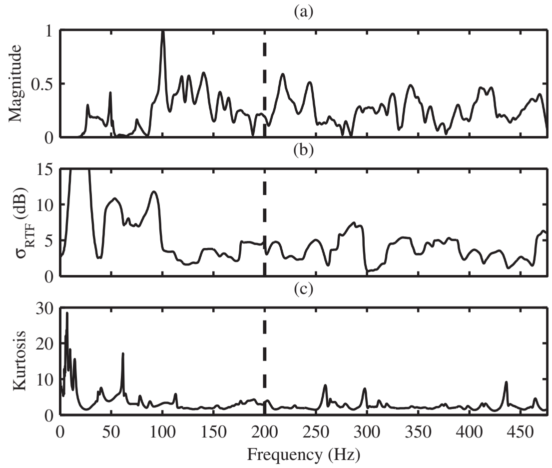

FIG. 7. (a) RTF magnitude in the office, (b) standard deviation of the logarithmic magnitude, and (c) kurtosis of the magnitude. The vertical dashed line indicates the theoretical Schroeder frequency.

whereas below it their values are significantly different. It can also be seen in Fig. 6(a) that the overall RTF magnitude is reduced below the Schroeder frequency, but this phenomenon was found to be caused by a reduced sound power output from the measurement loudspeaker. In Fig. 7, the results for the office (LG023 in Table I) show that the kurtosis fluctuates around the theoretical statistical value for the whole frequency bandwidth except at the very low end where the results are biased because less than the 80 frequency samples are available for the estimation of the statistics. The standard deviation exhibits a similar behavior except that it fluctuates around the theoretical value starting from 100 Hz. It is worth mentioning that analogous graphs from RTFs measured in other rooms showed that the estimated statistics often changes at a lower frequency than the theoretical Schroeder frequency.

B. Room volume from Schroeder frequency

The success of the room volume estimation method based on the Schroeder frequency is dependent on the accuracy with which the Schroeder frequency can be estimated from the RTF magnitude statistics and this poses the following challenges. The unbiased estimation of the magnitude statistics requires a width of the frequency window that incorporates more than a single magnitude peak or dip and consequently the width must be at least 20 Hz. As can be seen from Table I, the value of the Schroeder frequency in the larger rooms is of the same order and leaves no margin to estimate the magnitude statistics below it and many loudspeakers do also not produce sufficient sound power at these low frequencies. Further problems are that the magnitude statistics below the Schroeder frequency are not independent of source and receiver position and that a strong direct sound component has a bias effect on the magnitude statistics and may thus need to be excluded.

To ascertain whether the method can provide any affirmative results for the estimated room volume, its performance has been tested on all of the RTFs in nearly all of the available rooms. The experimentally observed Schroeder frequency was taken as the highest frequency sample where either the kurtosis was above 7 or the standard deviation was above 7 dB. These two values are slightly higher than their respective asymptotic values and were determined empirically by looking at graphs analogous to those shown in Figs. 6 and 7 from various RIRs. The numerical results in terms of mean and standard deviation between the receiver positions in each room are given in Table II and Fig. 8 shows a plot of the estimated versus the geometrical room volume.

TABLE II. Mean μ and relative standard deviation σ/μ of estimated volume V_{Sch} together with room volume V_{geo} from geometrical measurements. The numbers in parentheses following the name of the room indicate the fraction of receiver positions where a valid estimate was obtained.

Name	V_{geo} (m^3)	$\mu_{V_{\mathrm{Sch}}}$ (m^3)	$\sigma/\mu_{V_{\mathrm{Sch}}}$
Office (5/15)	60(80)	570	0.92
Listening room (0/10)	131	⋯	⋯
Lecture room A (61/143)	180(220)	533	0.96
Lecture room B (3/4)	550	15 000	0.11
McMordie Hall (5/9)	850	12 000	1.20
Harty Room (6/18)	1 150	7 000	1.78
Sonic Laboratory (12/30)	3 200	7 400	0.70
Whittla Hall (4/8)	8 400	38 000	0.96
Concert hall A (200/420)	19 000	20 000	0.94
Concert hall B (279/512)	24 000	906	0.92

From Table II, the method yields valid results at usually less than half of the receiver positions (valid means that the estimated Schroeder frequency is neither zero nor the maximum of the frequency range considered). Moreover, some of the estimates are either too large or too small, by almost two orders of magnitudes. Figure 8 shows that the correlation between estimated and true room volume is poor. Taking into account that it can already be guessed that the volume of the rooms encountered in practice is in the range between 5 and 40 000 m^3, it is concluded that this method yields neither consistent nor useful results.

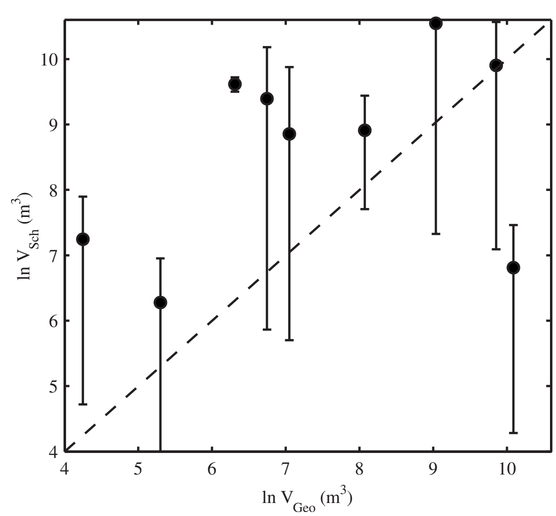

FIG. 8. The logarithm of the mean of V_{Sch} vs V_{Geo} with the error bars corresponding to the standard deviation between the receiver positions in each room. The dashed line indicates $V_{\mathrm{Sch}}=V_{\mathrm{Geo}}$.

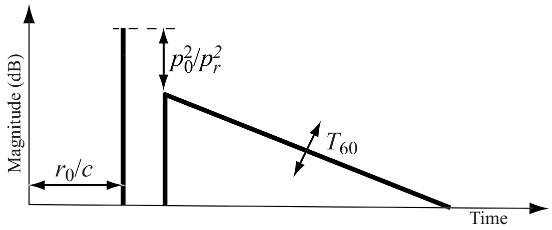

FIG. 9. Diagrammatic representation of the logarithmic magnitude in a RIR in terms of the arrival time of the direct sound r_0/c, the direct-to-reverberant ratio p_0^2/p_r^2, and the reverberation time T_{60}.

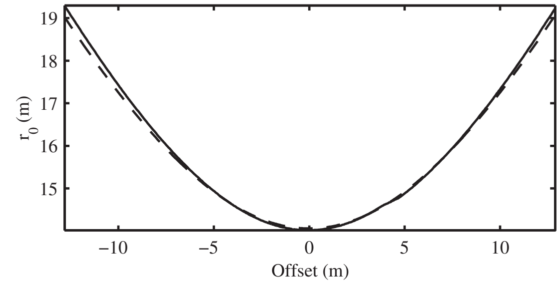

FIG. 10. Source-to-receiver distance r_0 across the width of concert hall B measured geometrically (---) and estimated acoustically (—).

V. DIFFUSE FIELD METHOD

From an energetic perspective, the logarithmic magnitude of a RIR can be described by the arrival time of the direct sound r_0/c, the reverberation time T_{60}, and the direct-to-reverberant ratio p_0^2/p_r^2 as illustrated in Fig. 9. If the source-to-receiver distance r_0 and the acoustic properties of the walls are kept constant when moving from a small room to a larger room, it seems logical to expect that the ratio p_0^2/p_r^2 would increase due to the decrease in reverberant energy density per unit volume. This observation is now formalized mathematically and forms the essence of the room volume estimation method based on diffuse field acoustics.

A. Theoretical basis

In a diffuse field, the mean square pressure of the reverberant sound field is given by[20]

$$\overline{p_r^2} = \rho_0 c W\left(\frac{T_{60}c}{6\ln(10)V}\right), \tag{13}$$

where $\rho_0 c$ is the specific acoustic impedance of air and W the sound power. The mean square reverberant pressure can be calculated from a RIR by summing all squared amplitudes therein but leaving out the direct sound component. The room volume can therefore be obtained from Eq. (13) but requires knowledge of W. This quantity is usually unknown. To circumvent this problem, Eq. (13) is accompanied by the equation for the mean square direct sound pressure given by[20]

$$\overline{p_0^2}(r_0) = \rho_0 c W\left(\frac{1}{4\pi r_0^2}\right), \tag{14}$$

where r_0 is the source-to-receiver distance. Again, the mean square pressure of the direct sound can be obtained from the squared amplitudes in the RIR.

When combining Eq. (14) with Eq. (13), the following expression for the room volume results:

$$V_{\text{classic}} = \frac{\overline{p_0^2}(r_0)}{\overline{p_r^2}}\frac{4\pi r_0^2 c T_{60}}{6\ln(10)}. \tag{15}$$

Alternatively, the reverberant pressure according to revised diffuse field theory is given by[1,2]

$$\overline{p_r^2} = \rho_0 c W\left(\frac{T_{60}c}{6\ln(10)V}\right)[e^{-r_0/c6\ln(10)}], \tag{16}$$

which leads to a slightly different alternative equation for the room volume given by

$$V_{\text{revised}} = \frac{\overline{p_0^2}(r_0)}{\overline{p_r^2}}\frac{4\pi r_0^2 c T_{60}}{6\ln(10)}[e^{-r_0/c6\ln(10)}]. \tag{17}$$

Moreover, in order to take the directivity of source and receiver into account, Eq. (14) has to be modified as follows:

$$\overline{p_0^2}(r_0) = Q_{\text{src}}(\mathbf{r}_0)Q_{\text{rec}}(-\mathbf{r}_0)\rho_0 c W\left(\frac{1}{4\pi r_0^2}\right), \tag{18}$$

with $Q_{\text{src/rec}}(\pm\mathbf{r}_0)$ the directivity factor of the source/receiver in the direction of the receiver/source, respectively. For the reverberant sound pressure, the directivity factors do not need to be included because they are by definition unity. Equation (18) does of course also alter Eqs. (15) and (17).

B. Calculation of the required parameters

The calculation of the room volume according to Eq. (15) or Eq. (17) requires parameter values for r_0, p_0^2, p_r^2, T_{60} and also in principle $Q_{\text{src}}(\mathbf{r}_0)$ and $Q_{\text{rec}}(-\mathbf{r}_0)$. The robust automatic estimation of each parameter from a RIR is now briefly discussed. Numerical results for each parameter are illustrated for RIRs measured across the width of concert hall B.

In a correctly measured RIR, the source-to-receiver distance r_0 can be estimated from the initial delay $\tau_0 = r_0/c$ of the direct sound arrival time. In the estimation procedure, τ_0 was assigned the value of the first time sample whose magnitude was less than 22 dB below the maximum magnitude in the entire RIR. This measure is necessary because the direct sound does not always correspond to the largest magnitude in the RIR. The very good agreement between geometrically measured and estimated r_0 is shown in Fig. 10. A general offset between the two curves is evident but this may also have been caused by incorrect geometrical measurements of r_0. More important, the estimated r_0 is consistent and shows no outliers.

For the estimation of the direct sound pressure p_0^2, the squared amplitudes in a time window, extending from −1 to 1.5 ms relative to the identified direct sound arrival time, are summed. For a better temporal resolution, the data have been resampled to 64 kHz. The estimated magnitude

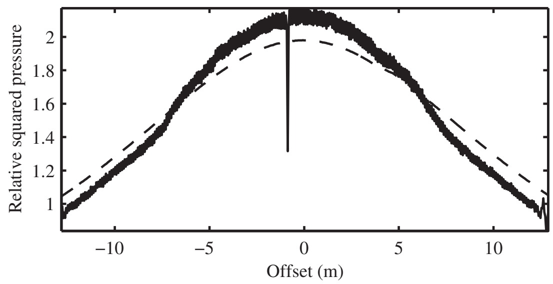

FIG. 11. Magnitude of the squared direct sound pressure p_0^2 across the width of concert hall B from the theoretical inverse square law (---) and estimated acoustically (—). The outlier slightly to the left of the zero offset is due to a measurement anomaly.

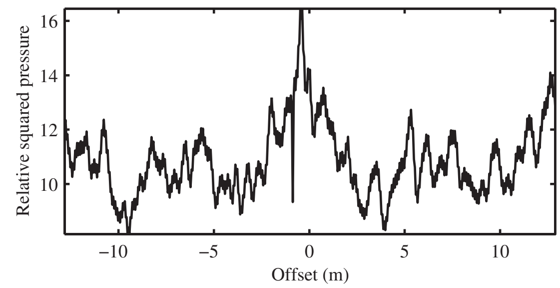

FIG. 13. Estimated magnitude of the squared reverberant sound pressure p_r^2 across the width of concert hall B. The relative standard deviation between receiver positions is 0.12.

and the theoretical behavior according to the inverse square law are shown in Fig. 11. Because of an arbitrary scaling value, the two curves have been rescaled for equal mean. The agreement between the two curves is quite good even though there is a systematic deviation at the central half of the receiver positions.

For the calculation of the reverberant pressure and the reverberation time, the method by Lundeby *et al.*[21] has been used to find the crossover point between the sound decay and the stationary noise floor and T_{60} is then obtained from a straight line fit to the logarithm of the energy decay curve obtained from Schroeder backwards integration.[22] Figure 12 shows the estimated value of T_{60} across the width of concert hall B, which demonstrates that the reverberation time does fluctuate with receiver position. The reverberant pressure has been obtained by summing the squared amplitudes in the RIRstarting from the end of the time window used for the direct sound pressure and stopping at the crossover point found by Lundeby's method. Figure 13 shows the estimated value of p_r^2 across the width of concert hall B and illustrates that this parameter does also fluctuate with receiver position.

Finally, no procedure has been found to estimate the directivity factor of the source and receiver from the given input data and it has thus been assumed that $Q_{src}(\mathbf{r}_0) = Q_{rec}(-\mathbf{r}_0) = 1$. It needs to be mentioned that it is possible to obtain the directivity data by other means such as from the manufacturer's datasheet or from measurements in an anechoic chamber. Because the directivity data are required for a arbitrary three-dimensional direction and for a wide frequency range, the present author has decided to not pursue this avenue further because it would severely limit the method's applicability to laboratory experiments.

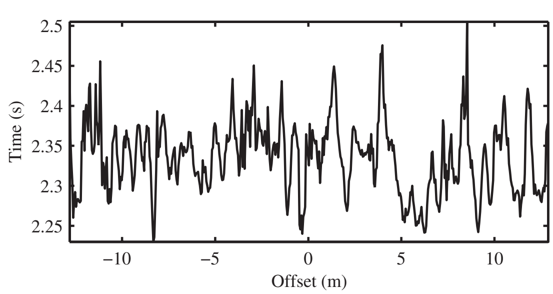

FIG. 12. Estimated reverberation time T_{60} across the width of concert hall B. The relative standard deviation between receiver positions is 0.02.

C. Results

The performance of the room volume estimation method has been evaluated with all RIRs measured in the twelve rooms listed in Table I. Initial results obtained in the different rooms showed that with the polyhedral loudspeaker the volume estimates were of the correct order, whereas with the studio monitor all estimates were by an approximate factor ten too large. This fact was attributed to the difference in directivity between the two sound sources. For the studio monitor, the directivity factor in the direction of the receiver significantly exceeds unity at higher frequencies. This results in an increased direct sound pressure and, when inserting into Eq. (15) or Eq. (17), also in an overestimated room volume. The problem has been circumvented by restricting the frequency bandwidth to frequencies where the source radiation is approximately omnidirectional. Experiments with several loudspeakers have shown that an upper frequency limit of 700 Hz is appropriate. At the same time, a lower frequency limit of 200 Hz, higher than the Schroeder frequency in most rooms, was also introduced. It will become evident in Sec. V E that the bandwidth is detrimental to the accuracy of the room volume estimate.

After incorporating these measures, Table III shows the numerical results for volume estimates $V_{classic}$ and $V_{revised}$ from classical and revised diffuse field theory in terms of mean and standard deviation between all N_R receiver positions in a room. Quoting the values for mean and standard deviation is based on the assumption that the data follow a normal distribution. Performing the Kolmogorov–Smirnov test[23] on the data from lecture hall A and concert hall B proved that this assumption is justified. Naturally, the result obtained with a low number of receiver positions are statistically less representative. For concert hall B, Fig. 14 illustrates the variation of $V_{revised}$ and $V_{classic}$ across the width of the room.

Excluding the results for concert hall A, the office, the Sonic Laboratory and the lavatory, $\mu_{V_{revised}}$ is generally within $\pm 50\%$ of V_{geo}. This good correlation is perhaps better visualized by plotting $V_{revised}$ against V_{geo} as shown in Fig. 15, which is also to be compared with Fig. 8 for the room volume estimation method based on the Schroeder frequency. Even though $\mu_{V_{revised}}$ is generally closer to V_{geo} than $\mu_{V_{classic}}$ is, in most rooms revised theory produces only marginally more consistent results than classical theory as evident by the comparable standard deviation. Without the four

TABLE III. Mean μ and relative standard deviation σ/μ of estimated volumes V_{classic} and V_{revised} together with room volume V_{geo} from geometrical measurements.

Name	V_{geo} (m^3)	$\mu_{V_{\text{classic}}}$ (m^3)	$\sigma/\mu_{V_{\text{classic}}}$	$\mu_{V_{\text{revised}}}$ (m^3)	$\sigma/\mu_{V_{\text{revised}}}$
Lavatory	5(7)	14	0.28	11	0.27
Office	60(80)	231	0.28	161	0.27
Listening room	131	206	0.44	130	0.39
MMedia room	150	303	0.33	213	0.36
Lecture hall A	180(220)	297	0.20	248	0.21
Lecture hall B	550	714	0.22	539	0.19
McMordie Hall	850	1 190	0.29	1 020	0.31
Harty Room	1 150	1 480	0.31	1 130	0.33
Sonic Laboratory	3 200	3 340	0.52	2 020	0.44
Whittla Hall	8 400	14 500	0.24	9 450	0.19
Concert hall A	19 000	5 460	0.25	4 540	0.23
Concert hall B	24 000	26 900	0.14	20 400	0.16

exceptions already mentioned, it can be concluded that the room volume estimation method based on diffuse field acoustics delivers consistent results for rooms ranging in size from 100 m^3 to 20 000 m^3, i.e., over nearly three orders of magnitude, and having substantially different acoustic characteristics.

D. Exception to the general result trend

The results presented in Table III have shown that concert hall A, the office, the Sonic Laboratory, and the lavatory are exceptions. In all of these rooms $\mu_{V_{\text{revised}}}$ differs by more than 50% from V_{geo}. In the office and the lavatory, the Schroeder frequency is equal to or larger than the lower frequency limit. Experiments revealed that in these two rooms $\mu_{V_{\text{revised}}}$ moves closer to V_{geo} when the frequency limit is increased. But it is shown further below that decreasing the frequency bandwidth has the detrimental effect of increasing the variance in the volume estimates.

In the Sonic Laboratory, $\mu_{V_{\text{revised}}}$ is slightly below -50% from V_{geo} and $\sigma/\mu_{V_{\text{revised}}}$ is the highest value of all rooms investigated. A particular design feature of this room is that the performers and the audience are located on a metal grid floor below which there is an undercroft of substantial volume. This raises the question whether the room should be treated as a single volume or two coupled volumes. Due to this issue, the homogeneity of the sound field is questionable.

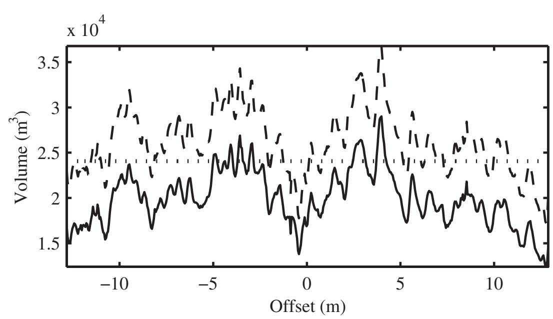

FIG. 14. Estimated room volumes V_{revised} (—) and V_{classic} (---) across the width of concert hall B. The true room volume is indicated by the dotted horizontal line.

In concert hall A, $\mu_{V_{\text{revised}}}$ is a factor of 4 too small and this result is consistent as shown by the low value for $\sigma/\mu_{V_{\text{revised}}}$. The estimates for r_0 and the T_{60} have been checked and found to yield plausible values, which leaves the squared pressure of the direct and reverberant sound. Cross referencing the ratio between the two with the same quantity in concert hall B of similar volume and complexity, it was found that the ratio is far too small for the same source-to-receiver distance. The most likely cause seems to be an uncharacteristically low magnitude of the direct sound but a further investigation into this anomaly was not possible because the data were not measured by the author and no further knowledge of the measurement setup was available. By comparing with the floor reflection, it was estimated that the squared direct sound pressure should theoretically be larger by an approximate factor three. This would increase the room volume estimate by the same factor and consequently bring it close to the value for V_{geo}.

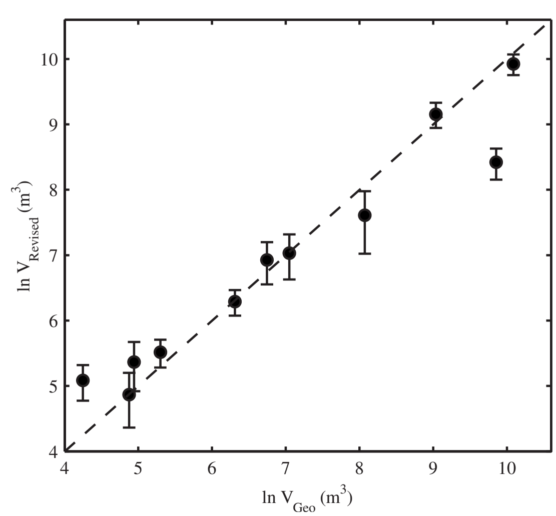

FIG. 15. The logarithm of the mean of V_{revised} vs V_{geo} with the error bars corresponding to the standard deviation between the receiver locations in each room. The dashed line indicates $V_{\text{revised}}=V_{\text{geo}}$.

E. Variance in the estimated parameters

It is now attempted to explain the variance of the room volume estimates in Table III in terms of both the theoretical and experimentally observed variance of the parameters involved in the calculation.

1. Variance in reverberant sound pressure and reverberation time

As mentioned in Sec. IV A, the standard deviation of the sound pressure for a single frequency is 5.57 dB. This value is decreased if the standard deviation is measured in a frequency band. Both Schroeder[24] and Lubman[25] considered this case and arrived at the same approximate equation for the standard deviation $\sigma_{\overline{p_r^2}}$ of the reverberant sound pressure level

$$\sigma_{\overline{p_r^2}} \approx \frac{5.57}{\sqrt{1+\frac{3.3BT_{60}}{13.8}}}\ (\mathrm{dB}), \tag{19}$$

where B is the equivalent bandwidth.[24] Equation (19) exhibits the correct asymptotic behavior in as much that for $B \to 0$ the single frequency value of 5.57 dB is obtained and for $B \to \infty$ the standard deviation tends to zero. Chiles and Barron[2] found the scatter of the sound pressure level in a scale and computer model to be larger than theoretically predicted by Eq. (19) but the agreement improves if only the late reverberant sound is considered.

An approximate expression for the relative standard deviation of the reverberation time can be obtained from formulas derived in Refs. 26 and 27 and is given by[28]

$$\frac{\sigma_{T_{30}}}{T_{30}} \approx \frac{0.59}{\sqrt{BT_{30}}}. \tag{20}$$

Here, T_{30} instead of T_{60} is used because the decay was measured over 30 dB and B is now the statistical bandwidth of the RIR or the bandpass filter. [29]

With the bandpass filter used in the current paper to attenuate the spectrum at frequencies below the Schroeder frequency and above the frequency where the source radiation is no longer omnidirectional, the equivalent or statistical bandwidth $B \approx 540$ Hz. Converting the dB value from Eq. (19) into linear units, the predicted relative standard deviation of the reverberant sound pressure for the rooms listed in Table III is in the range $\sigma_{\overline{p_r^2}}/\overline{p_r^2} \approx 0.06-0.21$ (depending on the reverberation time of the room in question). Similarly, Eq. (20) predicts that $\sigma_{T_D}/T_D \approx 0.02$ to 0.05. For concert hall B, the relevant experimental values, quoted in the captions of Fig. 13 and Fig. 12, are 0.12 and 0.02, respectively. Both of these values are in the respective range predicted by theory.

The values for the standard deviation of the reverberation time and reverberant sound pressure do account for part of the experimentally observed standard deviation in the estimated room volume and they do explain why the rooms with larger reverberation time mostly exhibit a smaller deviation in the room volume estimates. The combined theoretical standard deviation due to the T_{60} and $\overline{p_r^2}$ is given by[30]

$$\frac{\sigma_{V_{\mathrm{theory}}}}{V} = \sqrt{\frac{\sigma^2_{\overline{p_r^2}}}{(\overline{p_r^2})^2} + \frac{\sigma^2_{T_{60}}}{(\bar{T}_{60})^2} - 2\frac{\sigma_{\overline{p_r^2},T_{60}}}{\overline{p_r^2}\bar{T}_{60}}}, \tag{21}$$

where $\sigma_{\overline{p_r^2},T_{60}}$ is the covariance between the parameters $\overline{p_r^2}$ and T_{60} that will be nonzero because the variation in both parameters stems from the same physical wave phenomena. It has been found that the experimentally observed $\mu_{V_{\mathrm{revised}}}$ for any room listed in Table III is always underpredicted by the value given by Eq. (21). On the bandwidth issue, the volume estimation for the RIRs from concert hall B has been repeated with an upper frequency limit of 10 kHz and the result was a reduction of $\sigma/\mu_{V_{\mathrm{revised}}}$ to 0.10.

2. Variance in source-to-receiver distance and direct sound pressure

It was already concluded from Fig. 10 that the error in estimating the source-to-receiver distance from the initial time delay r_0/c is fairly small. The variation in the estimated squared direct sound pressure can be obtained by measuring the relative standard error between the two curves in Fig. 11; its is 0.06 for concert hall B. This error would need to be incorporated into Eq. (21) by an extra term under the square root. It seems reasonable to assume that this term is independent of the other terms due to the reverberation time and the reverberant sound pressure.

VI. CORRELATION BETWEEN ROOM VOLUME AND T_{60}

Apart from the directivity issue, the main drawback of the room volume estimation method presented in Sec. V is that it relies on correctly measured RIRs where the initial time delay can be used to calculate the source-to-receiver distance. In this section, it is investigated whether an approximate relationship between the room volume and the reverberation time can be used for a simplified estimation of the room volume.

Suppose it can be assumed that the surface area S of a room is related to the volume by $S=\beta V^{2/3}$. The minimum value of $\beta=6$ results for a cube and for a room with an aspect ratio of 2:1:1, $\beta=6.3$. In the following it is assumed that $\beta=6.4$ is an average representative value for the rooms encountered in practice. The Sabine equation for the reverberation time then reads

$$T_{60} \approx \frac{0.161V}{6.4\bar{\alpha}V^{2/3}} = \frac{0.161V^{1/3}}{6.4\bar{\alpha}}. \tag{22}$$

In Fig. 16, the average T_{60} per room in the 500 Hz octave band is plotted as a function of the natural logarithm of the room volume for all the rooms measured. The solid curve is the linear least-square fit of Eq. (22) for all the rooms indicated by the filled circles. It results in an average absorption value of $\bar{\alpha}=0.26$. The rooms excluded thus either have an uncharacteristically low or high value for $\bar{\alpha}$.

Alternatively, the dashed line in Fig. 16 is the linear least-square fit of a straight line through the plotted data. Its approximate equation is

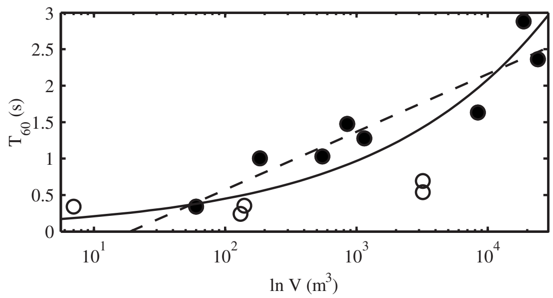

FIG. 16. T_{60} in the 500 Hz octave band vs the logarithm of the room volume V. The solid curve is the least-squares fit of Eq. (??) through all the closed circles. The open circles indicate the volume in rooms with particular acoustics. These are in order of increasing volume: Lavatory, office, the listening room, and the Sonic Laboratory. The dashed line is the linear least-squares fit $T_{60} \propto \ln V$.

$$T_{60} \approx 0.34 \ln(V) - 1. \tag{23}$$

Table IV shows the results after rearranging Eq. (23) to yield the volume from the reverberation time in the 500 Hz octave band (the 1 kHz octave band can similarly be used) for all rooms investigated. As expected, the room volume results show fairly large errors for those rooms which have not been included in the least-squares fit. Except for the Sonic Laboratory and the lavatory, the results are, however, mostly of the correct order of magnitude. A brief comparison between using Eq. (22) or Eq. (23) with $\bar{\alpha}=0.26$ for the room volume revealed that Eq. (23) yielded more accurate results in the Harty Room, McMordie Hall, and lecture room A whereas the results were slightly more inaccurate in concert hall A, the Sonic Laboratory and the Whittla Hall.

A. Source-to-receiver distance estimation

Due to the promising room volume results in Table IV, it seems logical to reconsider Eq. (15) and investigate to what extent it can be used to estimate the source-to-receiver distance. Inserting Eq. (23) into Eq. (15) and rearranging for r_0 yields

$$r'_0 \approx \sqrt{\frac{\overline{p_r^2}}{\overline{p_0^2}(r_0)} \frac{6 \ln(10)}{4\pi c T_{60}} e^{(T_{60}-1)/0.34}}. \tag{24}$$

The last two columns of Table IV show the mean and standard deviation of the relative difference between r'_0 and the true source-to-receiver distance r_0 in each room. In just under half of the rooms this procedure results in a mean error of approximately 20%. Larger errors of 70% result in rooms where the value for the volume was already inaccurate. The result for concert hall A is of course an exception because of issues already discussed in Sec. V D. Note that one reason for the small relative errors is the square root operation in Eq. (24).

VII. CONCLUSION

The estimation of the room volume from a given room impulse response has been investigated. The measurement data were obtained from several receiver positions in a total of 12 rooms of varied size and acoustic characteristics. The investigated methods were based on geometrical acoustics, eigenmode, and diffuse field models. It was found that the estimation method based on the temporal reflection density fails because of the difficulty in identifying reflections. The estimation method based on the Schroeder frequency was found to deliver inconsistent results because of the difficulty in experimentally determining the Schroeder frequency.

The estimation method based on diffuse field acoustics was found to deliver consistent results for room ranging in volume over almost three orders of magnitude. With the exceptions of the results in four rooms, the average of the estimated room volume between receiver positions was within $\pm 50\%$ of the true room volume with a standard deviation of approximately 0.25. The deviant results in the four excepted rooms were explained by a measurement error, a particularly high Schroeder frequency and a particular room design feature. A part of the experimentally observed standard deviation was explained by theoretical expressions for the standard deviation of the reverberant sound pressure and the reverberation time.

TABLE IV. Mean μ and relative standard deviation σ/μ of estimated volume from inverting Eq. (23) in the 500 Hz octave band. The last two columns express the error between the true r_0 and estimated source-to-receiver distance r'_0, where $\epsilon_{r_0}=|r_0-r'_0|/r_0$.

Name	V_{geo} (m^3)	μ_V (m^3)	σ/μ_V	$\mu_{\epsilon_{r_0}}$	$\sigma_{\epsilon_{r_0}}$
Lavatory	5(7)	51	0.10	0.78	0.31
Office	60(80)	50	0.04	0.41	0.08
Listening room	131	38	0.04	0.46	0.16
MMedia room	150	53	0.06	0.40	0.10
Lecture room A	180(220)	287	0.07	0.12	0.11
Lecture room B	550	378	0.13	0.14	0.19
McMordie Hall	850	1 370	0.08	0.22	0.21
Harty Room	1 150	784	0.16	0.13	0.19
Sonic Laboratory	3 200	120	0.23	0.71	0.12
Whittla Hall	8 400	2 180	0.10	0.50	0.08
Concert hall A	19 000	70 000	0.43	10.8	6.3
Concert hall B	24 000	22 560	0.09	0.13	0.08

Two drawbacks of this method are that it (i) relies on the presence of the initial time delay in the room impulse response for the estimation of the source-to-receiver distance and (ii) essentially assumes omnidirectional source and receiver. It was found that estimating the room volume from an approximate relationship with the reverberation time does not suffer from these drawbacks but does only provide accurate results for rooms whose absorption is neither uncharacteristically low nor high.

[1]M. Barron and L. J. Lee, "Energy relations in concert auditoriums. I," J. Acoust. Soc. Am. **84**, 618–628 (1988).
[2]S. Chiles and M. Barron, "Sound level distribution and scatter in proportionate spaces," J. Acoust. Soc. Am. **116**, 1585–1595 (2004).
[3]D. Cabrera, D. Jeong, H. J. Kwak, and J.-Y. Kim, "Auditory room size perception for modeled and measured rooms," in Proceedings of the 2005 Congres and Exposition on Noise Control Engineering, Rio, 2005.
[4]C. B. Pop and D. Cabrera, "Auditory room size perception for real rooms," in Proceedings of ACOUSTICS 2005, Australian Acoustics Society, Busselton, 2005.
[5]M. Kuster, D. de Vries, E. M. Hulsebos, and A. Gisolf, "Acoustic imaging in enclosed spaces: Analysis of room geometry modifications on the impulse response," J. Acoust. Soc. Am. **116**, 2126–2137 (2004).
[6]M. Monks, B. M. Oh, and J. Dorsey, "Audioptimization: Goal-based acoustic design," IEEE Comput. Graphics Appl. **20**, 76–91 (2000).
[7]A. T. Fürjes, F. Augusztinovicz, and E. Arató-Borsi, "A new method for the objective qualification of rooms," Acta Acust. **86**, 911–918 (2000).
[8]E. M. Hulsebos, "Auralization using wave field synthesis," Ph.D. thesis, Delft University of Technology, Delft, The Netherlands, 2004.
[9]J. van der Vorm, "Transform coding of audio impulse responses," Master's thesis, Delft University of Technology, Delft, The Netherlands, 2003.
[10]F. Li and T. J. Cox, "Speech transmission index from running speech: A neural network approach," J. Acoust. Soc. Am. **113**, 1999–2008 (2003).
[11]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. Am. **114**, 2877–2892 (2003).
[12]H. Kuttruff, *Room Acoustics*, 4th ed. (Spon, London, 2000), Chap. 4.
[13]J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950 (1979).
[14]E. Hecht, *Optics*, 2nd ed. (Addision-Wesley, Wokingham, 1987), Chap. 10.
[15]R. C. Gonzales and R. E. Woods, *Digital Image Processing*, 1st ed. (Addison-Wesley, Reading, Mass., 1992), Chap. 7.
[16]H. Kuttruff, *Room Acoustics*, 4th ed. (Spon, London, 2000), Chap. 3.
[17]M. Schroeder, "Statistical parameters of the frequency response curves of large rooms," J. Audio Eng. Soc. **35**, 299–306 (1987).
[18]M. Schroeder, "Normal frequency and excitation statistics in rooms: Model experiments with electric waves," J. Audio Eng. Soc. **35**, 307–316 (1987).
[19]Valid for the case where the receiver is sufficiently far away from the source in terms of direct-to-reverberant energy ratio.
[20]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982), Chap. 13.
[21]A. Lundeby, T. E. Vigran, H. Bietz, and M. Vorländer, "Uncertainties of measurements in room acoustics," Acustica **81**, 344–355 (1995).
[22]M. Schroeder, "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412 (1965).
[23]A. M. Mood, F. A. Graybill, and D. C. Boes, *Introduction to the theory of statistics*, McGraw–Hill Series in Probability and Statistics, 3rd ed. (McGraw–Hill, New York, 1974).
[24]M. R. Schroeder, "Effect of frequency and space averaging on the transmission responses of multimode media," J. Acoust. Soc. Am. **46**, 277–283 (1969).
[25]D. Lubman, "Fluctuations of sound with position in a reverberant room," J. Acoust. Soc. Am. **44**, 1491–1502 (1968).
[26]J. L. Davy, I. P. Dunn, and P. Dubout, "The variance of decay rates in reverberation rooms," Acustica **43**, 12–25 (1979).
[27]J. L. Davy, "The variance of impulse decays," Acustica **44**, 51–56 (1980).
[28]M. Barron, "Thoughts on the room acoustic enigma: The state of diffusion," *Proceedings of the Institute of Acoustics*, Oxford, 2005, Vol. 27.
[29]J. L. Davy and I. P. Dunn, "The statistical bandwidth of Butterworth filters," J. Sound Vib. **115**, 539–549 (1987).
[30]H. Ku, "Notes on the use of propagation of error formulas," J. Res. Natl. Bur. Stand., Sect. C **70**, 263–273 (1966).