Comparison of psychoacoustic-based reverberance parameters

Doheon Lee,[1,a)] Jasper van Dorp Schuitman,[2] Densil Cabrera,[3] Xiaojun Qiu,[1] and Ian Burnett[1]

[1]*Faculty of Engineering and Information Technology, University of Technology, Sydney, New South Wales 2007, Australia*
[2]*Sound Intelligence, Amersfoort, The Netherlands*
[3]*Sydney School of Architecture, Design and Planning, The University of Sydney, Sydney, New South Wales 2006, Australia*

This study compared psychoacoustic reverberance parameters to each other, as well as to reverberation time (RT) and early decay time (EDT) under various acoustic conditions. The psychoacoustic parameters were loudness-based RT (T_N), loudness-based EDT [EDT_N; Lee, Cabrera, and Martens, J. Acoust. Soc. Am. **131**, 1194–1205 (2012a)], and parameter for reverberance [P_{REV}; van Dorp Schuitman, de Vries, and Lindau., J. Acoust. Soc. Am. **133**, 1572–1585 (2013)]. For the comparisons, a wide range of sound pressure levels (SPLs) from 20 dB to 100 dB and RTs from 0.5 s to 5.0 s were evaluated, and two sets of subjective data from the previous studies were used for the cross-validation and comparison. Results of the comparisons show that the psychoacoustic reverberance parameters provided better matches to reverberance than RT and EDT; however, the performance of these psychoacoustic reverberance parameters varied with the SPL range, the type of audio sample, and the reverberation conditions. This study reveals that P_{REV} is the most relevant for estimating a relative change in reverberance between samples when the SPL range is small, while EDT_N is useful in estimating the absolute reverberance. This study also suggests the use of P_{REV} and EDT_N for speech and music samples, respectively.
© 2017 Acoustical Society of America. https://doi.org/10.1121/1.5005508

I. INTRODUCTION

Reverberation is one of the most important physical phenomena in room acoustics. To estimate reverberation, ISO 3382-1 (2009) recommends reverberation time (RT) and early decay time (EDT; hereafter, referred to as ISO reverberation parameters); however, these parameters do not represent the human perception of reverberation (hereafter, referred to as reverberance) well in some acoustic scenarios (e.g., Barron, 1988; Lokki *et al.*, 2012). This is, in part, due to auditory perception being a complex psychoacoustic process while the sound pressure level (SPL) decay values from which RT and EDT are derived represent very little of the transformation from sound to perception (Fastl and Zwicker, 2007).

To better estimate reverberance, Lee *et al.* (2012a) and van Dorp Schuitman *et al.* (2013) independently proposed psychoacoustic reverberance parameters. Although these parameters were shown to provide closer matches to reverberance than RT and EDT, they do not necessarily perform equally well because the parameters are derived from the output of different psychoacoustic models and emphasize different attributes of reverberance. In the present study, these psychoacoustic reverberance parameters are compared under various acoustic conditions to determine those that best represent reverberance over a range of acoustic conditions, and to better understand the strengths and limitations of these approaches.

The major differences between the psychoacoustic reverberance parameters are as follows: (1) Loudness-based RT (T_N) and loudness-based EDT (EDT_N) by Lee *et al.* (2012a) are based on the analysis of a room impulse response (RIR) in conjunction with the SPL, while parameter for reverberance (P_{REV}) by van Dorp Schuitman *et al.* (2013) is calculated directly from a running signal, e.g., music and/or speech; (2) T_N and EDT_N are based on the notion that reverberance comes from the perception of sound decay, while P_{REV} follows the theory from Griesinger (1997), proposing that reverberance is related to the absolute level of the room's reverberation and is independent of the level of the direct sound; and (3) the time-varying loudness model (TVL; Glasberg and Moore, 2002) and dynamic loudness model (Chalupper and Fastl, 2002) were used for T_N and EDT_N, while an auditory model based on the work by Breebaart (2001) and Breebaart *et al.* (2001) was used for P_{REV}.

On the basis of (2), the two sets of parameters are calculated in different ways. The calculation of T_N and EDT_N is similar to RT and EDT, i.e., a RIR is processed with a loudness model, and the decay time of its loudness decay function is calculated by analogy to RT and EDT. The slope of a loudness decay function becomes less steep as the SPL of a RIR increases (see Fig. 1 in Lee *et al.*, 2012a), so T_N and EDT_N values of a given RIR depend on the assumed listening level. For P_{REV}, a running signal is processed with the auditory model based on the work by Breebaart (Breebaart, 2001; Breebaart *et al.*, 2001), and the monaural outputs of the peripheral processor in the auditory model (which

a)Electronic mail: dosyd@hotmail.com

models outer/middle ear transfer function, basilar membrane, hair cells, and neural firing) are split into a foreground and background stream (as suggested by Griesinger, 1997; Rumsey, 2002; Mason et al., 2004). Then, P_{REV} is calculated as the average level of the background stream.

The differences in calculation method lead to distinct advantages and disadvantages of the parameters. As P_{REV} is based on a running signal, this parameter can account for the influence of sample on reverberance (Osses Vecchi et al., 2017), which can be an important contributor to reverberance (Teret et al., 2017). Furthermore, P_{REV} can estimate different characteristics of reverberance (such as stopped reverberance, running reverberance, and overall reverberance) (Morimoto and Asaoka, 2004) on the basis of analysing a selected part of a sample, and can be calculated both from audio recordings and during a live concert situation. However, for the same reason, a particular source-receiver pair of positions in a room can have different values of P_{REV} depending on which sample is selected and the processing thereof (e.g., inclusion or exclusion of silence at the end of the sample). In contrast, T_N and EDT_N are based on a RIR in conjunction with SPL, and beyond these the values are not influenced by the particular features of a running signal. Hence, these parameters are insensitive to some features of a sample that may affect reverberance. Furthermore, they cannot be calculated from live concert signals nor from audio recordings.

For the comparisons in this study, the parameters are calculated over a wide range of SPLs and RTs, and the cross-validation is performed with two sets of subjective data collected in the previous studies of the authors, i.e., Lee et al. (2012a) and van Dorp Schuitman et al. (2013). For the T_N and EDT_N calculations, the short-term loudness output of the TVL is used, since it approximates the momentary loudness perception of a binaural input signal (Moore and Glasberg, 2007). For the P_{REV} calculations, the auditory model based on the work by Breebaart (Breebaart, 2001; Breebaart et al., 2001) is used and the final decay of the music and speech is excluded from the analysis as in the study by van Dorp Schuitman (2011).

In Secs. II and III, the effect of SPL and T_{20} (i.e., RT with its evaluation range from -5 dB to -25 dB) on T_N, EDT_N, and P_{REV} are investigated, and then two sets of listening experiments performed by Lee et al. (2012a) and van Dorp Schuitman et al. (2013) are briefly described and the performance of the psychoacoustic reverberance parameters assessed. Finally, the main results are discussed and conclusions drawn.

II. EFFECT OF SPL AND T_{20} ON PSYCHOACOUSTIC PARAMETERS

In this section, the psychoacoustic reverberance parameters are calculated for various SPL and T_{20} values. T_N and EDT_N values are calculated from RIRs measured in a medium sized concert auditorium (Farina and Ayalon, 2003), and in a lecture theatre at the University of Sydney. The auditorium has 1200 seats with T_{20} of 1.84 s; the lecture theatre has 162 seats with T_{20} of 0.92 s. For P_{REV} calculations, these RIRs are convolved with an anechoic music excerpt from the

Denon Professional Test CD No. 2 (1988), i.e., Overture to *The Marriage of Figaro* by Mozart.

Figure 1 shows the effect of SPL on the psychoacoustic reverberance parameters, where L_{Aeq} (which is the power-averaged A-weighted SPL over a given time period) of the music sample ranges from 20 dB to 100 dB at 5 dB intervals. For T_N and EDT_N calculations, L_{AFmax} (which is the maximum A-weighted SPL with a "fast" temporal integration, i.e., using a 125 ms constant) of the RIRs is adjusted to these L_{Aeq} values. This range includes the SPLs for which T_N, EDT_N, and P_{REV} were validated in previous studies, e.g., from 60 dB to 80 dB in L_{Aeq} for T_N and EDT_N (see Table I, in Lee et al., 2012b), and from 47 dB to 79 dB in L_{eq} (which corresponds to approximately 43 dB–72 dB in L_{Aeq}) for P_{REV} (see Table I in van Dorp Schuitman et al., 2013). It should be noted that in the previous studies the parameters were not calculated at 5 dB intervals, and therefore the SPLs tested in Fig. 1 are not exactly the same as those tested in the previous studies.

As shown in Fig. 1, the psychoacoustic reverberance parameters increase with the SPL, which reflects a positive relationship between reverberance and listening level (Lee and Cabrera, 2010; Hase et al., 2000). Figure 1 shows that P_{REV} values rise more steeply with SPLs from 60 dB to 100 dB than at lower SPLs. Compared to P_{REV}, T_N, and EDT_N values increase more gradually over the entire SPL range, and they stop increasing or start decreasing at around 90 dB.

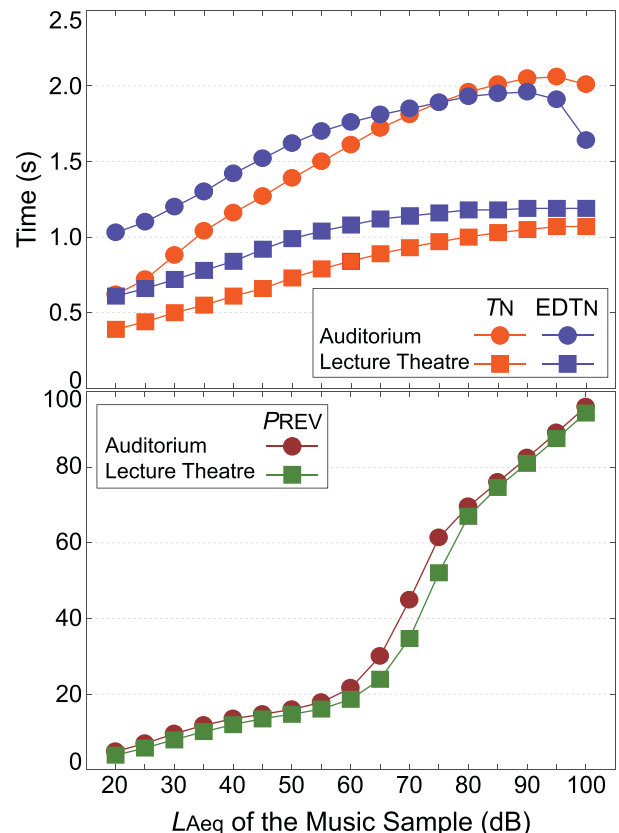Figure 2 shows the variations of the psychoacoustic reverberance parameters for T_{20} ranging from 0.5 s to 5.0 s

FIG. 1. (Color online) T_N, EDT_N, and P_{REV} as functions of the SPL from 20 dB to 100 dB with T_{20} being 1.84 (Auditorium) and 0.92 s (Lecture Theatre). The unit of T_N and EDT_N is time in seconds, and P_{REV} is a unitless parameter.

J. Acoust. Soc. Am. **142** (4), October 2017

Lee et al. 1833

TABLE I. An overview of the properties of experiments A.

	S1	S2	M1	M2	M3	L1	L2	L3
Source-receiver distance (m)	12	24	10	19	31	20.5	30	48
EDT_{mid} (s)	1.89	1.98	1.83	1.77	2.00	2.44	2.25	2.38
$T_{20,mid}$ (s)	2.06	2.07	2.01	2.03	2.17	2.66	2.60	2.53
Experiment A.1, L_{Aeq} (dB)	76.0	75.6	75.5	73.7	72.4	71.3	70.7	65.1
Experiment A.2, L_{Aeq} (dB)	77.5	76.1	76.8	74.6	73.5	72.1	71.1	65.2

with 0.5 s intervals. For this comparison, T_{20} of the auditorium RIR used for Fig. 1 was changed by modifying its decay envelope in the method suggested by Cabrera *et al.* (2011). To summarise the method, the RIR was processed with octave-band filters from 32.5 Hz to 16 kHz, and noise floors that come after the reverberation decays of the filtered RIRs were decayed at the same rate of the corresponding reverberation decay envelope. Then, the noise-treated octave-band RIRs were multiplied by exponential functions

FIG. 2. (Color online) T_N, EDT_N, and P_{REV} as functions of T_{20} from 0.5 s to 5.0 s, with the SPLs being 40 dB, 60 dB, and 80 dB. The unit of T_N and EDT_N is time in seconds, and P_{REV} is a unit-less parameter.

to achieve the desired octave-band RTs and combined into a single RIR.

In Fig. 2, T_N and EDT_N values increase almost linearly with T_{20}, and the effect of T_{20} becomes stronger as the SPL increases. For T_N at 60 dB and 80 dB and EDT_N at 80 dB, these parameters increase at a rate similar to that of T_{20}. In contrast, P_{REV} is scarcely affected by T_{20}, but is strongly affected by the SPL. For example, a tenfold increase in T_{20} (from 0.5 s to 5.0 s) increases P_{REV} at 80 dB by approximately 2% only (from 69.41 to 70.88). However, when the SPL increases from 40 dB to 80 dB, P_{REV} at T_{20} of 1.5 s is increased by more than five times (from 12.86 to 70.19).

It should be noted that, in previous studies, T_N and EDT_N have been validated for T_{20} from 1.0 s to 3.0 s (Lee *et al.*, 2012b), and P_{REV} has been validated for T_{20} from 0.02 s to 10.12 s (van Dorp Schuitman *et al.*, 2013). However, the tested T_{20} values in the previous studies were not at 0.5 s intervals. For this reason, T_{20} values in Fig. 2 are not exactly the same as those tested in the previous studies, although they are in the same range.
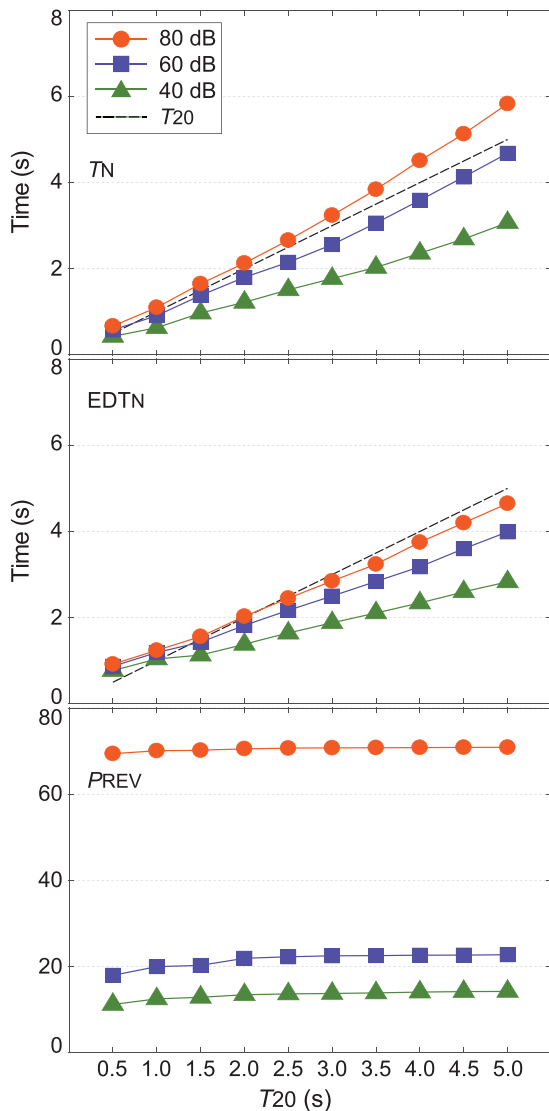
III. CROSS VALIDATION

In this section, cross-validation is performed with the subjective data collected in the studies of Lee *et al.* (2012a) and van Dorp Schuitman *et al.* (2013) in order to discover the acoustic conditions in which each psychoacoustic reverberance parameter performs best (either by yielding the lowest coefficient of variation for reverberance-matched samples or the highest correlation coefficient with the collected subject responses).

A. Matching reverberance experiments

1. Experiments

Lee *et al.* (2012a) performed listening experiments with two types of samples: orchestral music and a tenor solo voice singing an operatic excerpt. Eight RIRs measured by Farina and Ayalon (2003) in three auditoria were convolved with the anechoic recordings to generate the base stimuli. The detailed acoustic conditions are given in Table I, and are labeled to indicate the auditorium size [i.e., small (S), medium (M), and large (L)] and a number ordinally indicating source-to-receiver distance. It should be noted that the values in Table I are from RIRs with the noise floor treatment (in the same way applied for Fig. 2), which was necessary to avoid auralization artefacts when RT was increased as part of the listening test [see Eq. (1)]. The signal-to-noise ratios of the original RIRs were higher than 60 dB, and the noise floor treatment was applied from the point that is 10 dB above the noise floor. The treatment changed EDT_{mid} and $T_{20,mid}$ of the RIRs by less than 0.01 s (hereafter, the subscript "mid" refers to an average of its value at 500 Hz and 1 kHz in octave bands, and the subscript "oct" refers to an average of its values from 125 Hz to 4 kHz in octave bands).

In this paper, the experiment with the instrumental music samples is referred to as experiment A.1 and the experiment with a tenor voice is referred to as experiment

A.2. Note that Lee *et al.* (2012a) investigated the effect of SPL on reverberance by adding ± 5 dB gains to the convolution products, so the actual SPLs tested in that study were approximately from 60 dB to 82 dB in L_{Aeq}.

The task was to match the reverberance of 24 comparison samples (8 RIRs $\times$ 3 gains of -5, 0, and $+5$ dB) to the reverberance of one reference sample. The reference sample in each experiment was the respective anechoic recording convolved with the RIR identified as M1 (Table I). The subjects listened to 24 pairs of samples that consisted of the reference sample and one of the 24 comparison samples, and pressed "more" or "less" buttons on a MATLAB (MathWorks, Natick, MA) graphical user interface to adjust the reverberation decay of RIRs, which was then convolved with the anechoic samples for listening. Pressing the buttons incremented or decremented d in Eq. (1), where the $p(t)$ is sound pressure of a RIR as a function of time, t is time in seconds, and $p'(t)$ is sound pressure of the RIR as a function of time after the reverberant decay has been adjusted.

A unit step of d corresponds to approximately a 4% change in EDT_{mid}, which is similar to the 5% just noticeable difference (JND) of reverberance specified in ISO 3382-1 (2009). The initial value of d for the comparison samples was randomly chosen from -7 to 7, so that the samples were played with randomized reverberance at first. It should be noted that Eq. (1) changes the energy of a RIR, and this is compensated by multiplying ΔL_E in Eq. (2) to RIRs before the convolution process. That way, an undesired change in L_{Aeq} due to the change in reverberation decay of a RIR is avoided:

$$p'(t) = p(t)\exp\left(\frac{(-3 + (3 \times 1.04^d)t)}{1.04^d}\right), \qquad (1)$$

$$\Delta L_E = 10 \, \text{Log}_{10}\left(\frac{\int p'^2(t)}{\int p^2(t)}\right). \qquad (2)$$

The samples were presented via circumaural headphones (Sennheiser HD600, Sennheiser, Hanover) in an anechoic chamber at the University of Sydney in Australia for experiment A.1 and in a listening booth at Tohoku University in Japan for experiment A.2. Twenty subjects and 15 subjects took part on a volunteer basis in experiments A.1 and A.2, respectively. None of the subjects self-reported any hearing loss. As one of the comparison samples was physically identical to the reference sample, subjects who mismatched the reverberance of this pair by more than two times the JND of the reverberance (i.e., a 10% change in EDT_{mid}) were considered insensitive to the task and were excluded from analyses. To assess this, EDT_{mid} of the reference RIR is compared with that of the comparison RIR (which is physically identical to the reference RIR) incorporating the d-adjustment by each subject. The number of subjects excluded from the analyses was 4 in each experiment (note that they are not the same subjects), and therefore responses from 16 and 11 subjects were analysed further for experiments A.1 and A.2, respectively.

2. Results

The subject responses represented by d in Eq. (1) are given in Fig. 3. In Fig. 3, each box represents 25th, 50th, and 75th percentiles with whiskers extending to the most extreme d values that are not outliers. Outliers marked with a red cross are d values beyond 1.5 times the interquartile range from 25th and 75th percentiles. The interquartile range is defined as the difference between 25th percentile and 75th percentile. As shown in Fig. 3, M1 has median (i.e., 50th percentile) values close to zero (i.e., -0.5 and 0 in experiments A.1 and A.2, respectively), indicating that the subject matched reverberance correctly for the two physically identical samples. According to a three-way analysis of variance (ANOVA), the subject responses in d are significantly affected by the additional SPL changes of ± 5 dB [$F(2,614) = 33.81$, $p < 0.01$], the RIRs [$F(7,614) = 10.66$, $p < 0.01$], and the experiment number [$F(1,614) = 7.72$, $p < 0.01$]. However, there is no significant interaction effect between these independent variables, i.e., [$F(7,614) = 1.02$, $p = 0.414$] between the RIR and the experiment number; [$F(2,614) = 0.15$, $p = 0.857$] between the gain and the experiment number; and [$F(14,614) = 0.7$, $p = 0.773$] between the RIR and the gain.

The T_N, EDT_N, and the ISO reverberation parameters were derived from the reverberance-matched RIRs (of which reverberation decays were adjusted using the averaged d values). For P_{REV} calculations, these RIRs were convolved with the anechoic recordings of the orchestral music and the tenor voice singing an operatic excerpt.

The coefficient of variation between the 24 reverberance-adjusted samples was calculated for each parameter, and results are shown in Fig. 4. The coefficient of variation is the standard deviation divided by mean and, hence, a mean-related bias in the standard deviation is eliminated. As reverberance of the 24 comparison samples was subjectively matched to the same reference reverberance, the coefficient of variation between the reverberance-adjusted samples is ideally
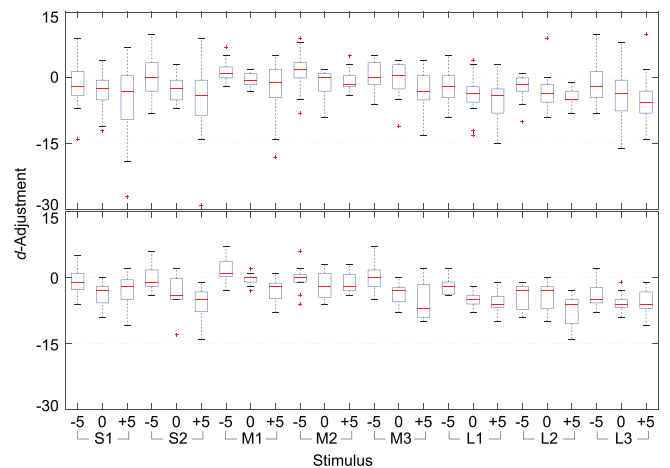

FIG. 3. (Color online) d-adjustments made by the subjects for matching reverberance [see Eq. (1)]. The numbers in the x axis represent the additional SPL changes of $\pm$ 5 dB applied to the samples. Boxes represent 25th, 50th (red line), and 75th percentiles with whiskers extending to 1.5 times the interquartile range from 25th and 75th percentiles. Red crosses are outliers that are d values beyond the whiskers.
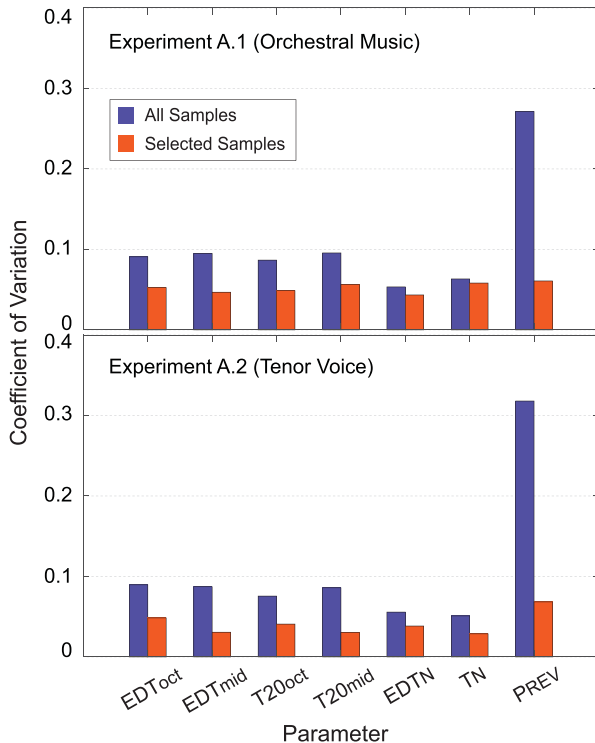
FIG. 4. (Color online) Values of the coefficient of variation for each parameter calculated from 24 reverberance-adjusted samples. The upper figure is for experiment A.1 (the orchestral music samples) and the lower figure is for experiment A.2 (the tenor voice singing an opera excerpt). The selected samples are S1 (-5 dB), S2 (-5 dB), M1 (-5 dB), L1 (0 dB), L2 (0 dB), and L3 ($+5$ dB), over which SPL range is less than 1 dB. The coefficient of variation of an ideal parameter would be zero for the reverberance-adjusted samples.

zero. One set of results is obtained from analysing all the samples, and the other set is obtained from analysing only six RIRs: S1 (-5 dB), S2 (-5 dB), M1 (-5 dB), L1 (0 dB), L2 (0 dB), and L3 ($+5$ dB), over which SPL difference is less than 1 dB. Hence, the effect of SPL is eliminated in the analysis of the latter set.

For the analysis of all samples, T_N and EDT_N have the lowest coefficients of variation, and P_{REV} yields the highest coefficient of variation. The significance of differences between coefficients of variation is examined with a two-sample f-test. It should be noted that this test is originally for testing the significance of difference between variances or between standard deviations. However, in this study, a two-sample f-test was executed on values of each parameter divided by its mean (which is also necessary because the P_{REV} values are different in scale to the other values), so that the result shows the significance of difference between coefficients of variation because the coefficient of variation is the standard deviation divided by mean.

According to the result of a two-sample f-test, in experiment A.1 the coefficient of variation of EDT_N is significantly lower than the other parameters ($p < 0.01$ for all pairs of parameters with EDT_N) except between EDT_N and T_N [$F(1,23) = 0.71$, $p = 0.425$]. In experiment A.2, the coefficients of variation of EDT_N and T_N are significantly lower than the other parameters ($p < 0.05$ for all pairs of parameters with EDT_N and T_N) except between EDT_N and T_N

[$F(1,23) = 1.21$, $p = 0.651$], between EDT_N and $T_{20,oct}$ [$F(1,23) = 1.85$, $p = 0.148$], and between T_N and $T_{20,oct}$ [$F(1,23) = 2.24$, $p = 0.059$]. For P_{REV}, its coefficients of variation are significantly higher than the other parameters ($p < 0.01$ for all pairs of parameters with P_{REV}) both in experiments A.1 and A.2. On the basis of this result, Table II tabulates the psychoacoustic reverberance parameters that are significantly better matches to reverberance than each ISO parameter. Table II shows that EDT_N is a better match to reverberance in experiments A.1 and A.2, while T_N is a better match only in experiment A.2.

For the analysis of the selected samples (for which the range of SPL across the samples is less than 1 dB), in the two experiments all the tested parameters are similarly matched to the subject responses. Results of a two-sample f-test show that the differences in coefficients of variation between the parameters are not statistically significant ($p > 0.05$ for all pairs of the parameters). The interesting result here is that P_{REV} is not an outlier for these samples, which were very similar in SPL.

B. Reverberance evaluation experiments

1. Experiments

van Dorp Schuitman *et al.* (2013) conducted four experiments in which subjects listened to four sets of binaural audio samples and rated four acoustic qualities, namely, reverberance, clarity, apparent source width, and listener envelopment, on a range from "very low" to "very high." Only the responses for reverberance are used in the present study. Each set of binaural recordings represents different acoustic conditions as listed in Table III. For the samples, an anechoic solo cello recording and anechoic speech were convolved with the four sets of measured or simulated binaural room impulse responses (BRIRs).

The experiments were conducted with a double-blind task, following a so-called "mixed procedure" method proposed by Chevret and Parizet (2007), which is a mix between a paired comparison and a direct evaluation method. Using this method, the subjects are allowed to apply direct rating to the samples using a slider on the screen and then the collected subject responses can be sorted from the highest to lowest rating, allowing for paired comparisons by fine-tuning the ratings. Chevret and Parizet applied this method to assess the perceived quality of car door closing sounds and showed that this method yields the same quality of responses as a paired comparison test, but with much shorter testing times. The same test procedure has been followed in experiments B, where subjects were asked to rate

TABLE II. Psychoacoustic reverberance parameters that are significantly better matches to reverberance than each ISO parameter in experiments A.1 and A.2 for all samples ($p < 0.05$).

	EDT_{oct}	EDT_{mid}	$T_{20,oct}$	$T_{20,mid}$
A.1	EDT_N	EDT_N	EDT_N	EDT_N
A.2	EDT_N, T_N	EDT_N, T_N	None	EDT_N, T_N

1836   J. Acoust. Soc. Am. **142** (4), October 2017

Lee *et al.*

TABLE III. An overview of the properties of experiments B.

Experiment number	Number of rooms	Room type	Number of subjects	Loudness normalized	Usage
B.1	9	Virtual (realistic)	5	Yes	Validation
B.2	8	Virtual (unrealistic)	5	Yes	Validation
B.3	10	Real	15	No	Training
B.4	10	Real	15	Yes	Validation

room acoustical qualities like "reverberance" using this mixed procedure.

As seen in Table III, experiments B.1 and B.2 included "virtual" rooms, for which the BRIRs were simulated using an acoustic shoebox model (van Dorp Schuitman, 2011). Five expert subjects participated in these experiments. They were from the acoustics department at TU Delft with in-depth knowledge about the room acoustical parameters and had experience in assessing those parameters. All subjects reported normal hearing. Before the start of the experiments, the subjects received instructions (including audio examples) explaining reverberance.

The main difference between the two tests is that for experiment B.1 "realistic" rooms were chosen with a large spread in RT, whereas rooms for experiment B.2 have more "non-realistic" properties in terms of dimensions, shape, and spatial distribution of absorption (e.g., one of the tested rooms had a long RT but side walls that were fully absorbing). In both tests, the samples were normalized to the same estimated loudness using the Replaygain 1.0 algorithm (Robinson, 2001).

For experiments B.3 and B.4, the BRIRs were measured using the ITA dummy head (Schmitz, 1995) for convolutions with the anechoic samples, and RIRs were measured using an omnidirectional microphone (Type 4134, Brüel and Kjær, Nærum) for the calculation of ISO parameters. The only difference between the two experiments is that for experiment B.4 all samples were normalized, whereas for experiment B.3 the samples retained their original loudness differences. Tables IV–VII list $T_{20,mid}$, EDT_{mid} and L_{Aeq} (taking the transfer function of the headphones into account) of the samples used in experiments B.1–B.4. Fifteen subjects participated in experiments B.3 and B.4. They consisted mostly of students with mixed musical experiences and preferences, and received the same instructions (including the audio examples) explaining reverberance as in experiments B.1 and B.2. None of the subjects reported hearing problems.

2. Results

The performances of the parameters in experiments B are compared by calculating correlation coefficients between the parameters and the subject responses, as in the study of van Dorp Schuitman et al. (2013). The correlation coefficient indicates the strength and direction of the linear relationship between two factors, and its value ranges from -1 to $+1$ (Privitera, 2015). Therefore, an ideal reverberance parameter would yield a correlation coefficient of $r = 1$. As each subject may have rated "very low" and "very high" differently on the continuous scale, the subject responses were normalized according to ITU-R BS.1284-1 (ITU-R, 2003) to compensate for variations in interpretation of the scale

$$z_i = \frac{x_i - \bar{x}_i}{\sigma_i} \sigma + \bar{x}, \qquad (3)$$

where z_i is the normalized results for subject i, x_i are the results for subject i, $\bar{x}_i$ is the mean result for this subject, and σ_i is the standard deviation. $\bar{x}$ and σ are the mean and the standard deviation, respectively, for all subjects.

Tables VIII and IX tabulate values of the correlation coefficient between the normalized subject responses (hereafter, subject responses) and the reverberance parameters for the cello samples and the speech samples, respectively. All the r-values in Tables VIII and IX are statistically significant ($p < 0.05$), except those showing negative correlations in Table IX. The highest correlation observed in each experiment is in boldface.

In Tables VIII and IX, P_{REV} is most highly correlated with the subject responses in experiments B.1, B.2 (only for the speech samples), and B.4. The significance of the difference between correlation coefficients is assessed with a Meng's z-test (Meng et al., 1991), which is a z-test on the basis of the number of samples and the three sets of correlation coefficients (e.g., r_{xz}, r_{yz}, and r_{xy}, where x and y are parameter values and z signifies the subject responses). The result shows that the differences in correlation coefficient between P_{REV} and the other parameters are statistically significant ($z > 1.5$, $p < 0.01$ for all the possible pairs of parameters with P_{REV}). By contrast, the highest correlation coefficient of EDT_N in experiments B.3 is only statistically significant when it compares with P_{REV} for the cello samples ($n = 10$, $z = 1.87$, $p = 0.031$).

TABLE IV. $T_{20,mid}$, EDT_{mid}, and L_{Aeq} values of the samples used in experiment B.1.

Room number	1	2	3	4	5	6	7	8	9
$T_{20,mid}$ (s)	0.01	0.33	0.72	0.73	1.81	1.91	1.40	2.02	6.92
EDT_{mid} (s)	0.01	0.08	0.85	0.83	2.05	1.73	1.39	2.14	7.01
L_{Aeq}, cello (dB)	68.9	68.3	68.7	69.0	68.7	68.9	69.2	69.7	68.5
L_{Aeq}, speech (dB)	64.7	65.2	65.1	65.1	65.7	65.3	64.9	65.8	66.3

TABLE V. $T_{20,mid}$, EDT_{mid}, and L_{Aeq} values of the samples used in experiment B.2.

Room number	1	2	3	4	5	6	7	8
$T_{20,mid}$ (s)	1.27	1.21	1.25	1.75	1.77	1.82	1.72	1.98
EDT_{mid} (s)	1.42	1.39	1.49	2.06	1.83	1.72	1.87	1.86
L_{Aeq}, cello (dB)	69.5	69.3	69.2	69.9	69.3	68.8	69.2	69.1
L_{Aeq}, speech (dB)	66.2	65.6	66.2	65.7	66.1	65.2	66.2	66.1

J. Acoust. Soc. Am. **142** (4), October 2017

Lee et al. 1837

TABLE VI. $T_{20,mid}$, EDT_{mid}, and L_{Aeq} values of the samples used in experiment B.3.

Room number	1	2	3	4	5	6	7	8	9	10
$T_{20,mid}$ (s)	0.02	0.21	0.76	1.02	1.21	1.67	2.29	3.94	4.81	10.12
EDT_{mid} (s)	0.02	0.13	0.79	0.98	0.88	1.44	1.12	4.83	2.11	9.84
L_{Aeq}, cello (dB)	45.9	60.5	60.0	62.9	47.8	52.5	48.3	60.4	47.2	72.2
L_{Aeq}, speech (dB)	42.8	59.4	58.5	61.2	45.8	48.4	46.0	58.8	44.2	71.2

Based on the results of a Meng's z-test, Table X tabulates psychoacoustic reverberance parameters that are significantly more highly correlated with reverberance than each ISO parameter. As shown in Table X, P_{REV} is a significantly better match to reverberance than all the ISO parameters in experiments B.1, B.2 (only for the speech samples), and B.4. EDT_N is significantly better correlated to reverberance than only some ISO parameters in experiments B.1 and B.4, but T_N is not a significantly better match to reverberance than any ISO parameter in experiments B.

With respect to the effect of sample type (cello/speech), results of a paired-sample t-test show that the sample type does not significantly affect the tested parameters and the subject responses, except P_{REV} in experiment B.1 [$t(16) = -2.39$, $p = 0.030$] and in experiment B.2 [$t(14) = -13.94$, $p < 0.01$].

IV. DISCUSSION

This study examines the performance of psychoacoustic reverberance parameters in various listening conditions, for reverberance-matched samples (experiments A), and for samples from which subjective scale values have been derived (experiments B). In the first case, ideally there should be no variation in reverberance between the reverberance-matched samples, and so a small coefficient of variation is sought. In the second case, performance is indicated by correlation between the reverberance parameter and the subjective scale values. A two-factor f-test and a Meng's z-test were executed to investigate the significance of difference between coefficients of variation in experiments A, and between correlation coefficients in experiments B, respectively.

The two parameters EDT_N and T_N were proposed because in the experiments by Lee and Cabrera (2010) and Lee et al. (2012a) they both seemed to be plausible approaches, and it was not clear which was better as a reverberance parameter. Since then, Lachenmayr (2016) compared these in a subjective test using a room with electro-acoustically augmented reverberation, finding that EDT_N provided better performance than T_N (and also better than other parameters tested). Results of the current study (especially from experiments B) provide

TABLE VII. $T_{20,mid}$, EDT_{mid}, and L_{Aeq} values of the samples used in experiment B.4.

Room number	1	2	3	4	5	6	7	8	9	10
$T_{20,mid}$ (s)	0.02	0.21	0.76	1.02	1.21	1.67	2.29	3.94	4.81	10.12
EDT_{mid} (s)	0.02	0.13	0.79	0.98	0.88	1.44	1.12	4.83	2.11	9.84
L_{Aeq}, cello (dB)	69.3	67.1	68.8	68.1	69.5	66.0	64.2	65.2	64.5	65.7
L_{Aeq}, speech (dB)	70.1	69.3	69.7	69.7	70.2	65.9	65.8	66.3	64.9	67.7

TABLE VIII. The correlation coefficients between the tested reverberance parameters and the normalized subject responses collected from experiments B with cello samples. The highest correlation observed in each experiment is in boldface.

	EDT_{oct}	EDT_{mid}	$T_{20,oct}$	$T_{20,mid}$	EDT_N	T_N	P_{REV}
B.1	0.88	0.85	0.85	0.84	0.89	0.85	**0.99**
B.2	0.84	0.83	**0.97**	0.96	0.32	0.93	0.87
B.3	0.88	0.88	0.86	0.86	**0.93**	0.84	0.77
B.4	0.78	0.78	0.78	0.79	0.87	0.81	**0.95**

further support for the use of EDT_N (rather than T_N) as a reverberance parameter.

For P_{REV}, the best performance is observed in experiments B.1 and B.4, which tested a small range of SPL (less than 5 dB due to the SPL adjustment for constant loudness) and large range of T_{20} (more than 5.0 s). Considering that the samples in experiment B.3 are the same as those in experiment B.4, apart from their SPLs, P_{REV} is evidently overly sensitive to variation in SPL. This high sensitivity to the SPL is, in part, because the non-linear adaptation loops in the auditory model yield overshoots that are sometimes too large. The high sensitivity to the SPL can be seen in Fig. 2, and is also the main reason why P_{REV} yields lower values of the coefficient of variation in experiments A for the selected samples (for which the SPL range is less than 1 dB) than for all the samples (in which SPL range exceeded 20). Hence, P_{REV} appears to be most useful for loudness-matched samples, or at least for samples across which the SPL varies little. It should be noted that the ISO parameters also yield lower values of the coefficient of variation for the selected samples in experiments A. However, unlike P_{REV}, this is because they do not consider the SPL influence on reverberance, and therefore the elimination of this influence by analysing the selected samples results in values closer matching reverberance.

P_{REV} also appears to be particularly useful for evaluating reverberance of speech, and it significantly better correlates with the subjective responses than any of the conventional parameters in all of the speech experiments, except in experiment B.3 (see Table X). In Table IX, EDT_N ($r = 0.84$) achieves a higher correlation coefficient than P_{REV} ($r = 0.81$), but those correlations are not significantly different according to a Meng's z-test ($n = 10$, $z = 0.41$, $p = 0.341$).

The most remarkable performance from P_{REV} is in experiment B.2, where, for speech samples, it achieves a

TABLE IX. The correlation coefficients between the tested reverberance parameters and the normalized subject responses collected from experiments B with speech samples. All the r-values in the table are significant at a confidence level of 95% ($p < 0.05$), except the underscored values. The highest correlation observed in each experiment is in boldface.

	EDT_{oct}	EDT_{mid}	$T_{20,oct}$	$T_{20,mid}$	EDT_N	T_N	P_{REV}
B.1	0.83	0.80	0.79	0.79	0.85	0.79	**0.96**
B.2	−0.60	−0.55	−0.26	−0.30	−0.31	−0.29	**0.88**
B.3	0.76	0.76	0.74	0.74	**0.84**	0.70	0.81
B.4	0.73	0.73	0.72	0.73	0.85	0.75	**0.96**

TABLE X. Psychoacoustic reverberance parameters that are significantly more highly correlated with reverberance in experiments B ($p < 0.05$).

	EDT_{oct}	EDT_{mid}	$T_{20,oct}$	$T_{20,mid}$
B.1 cello	P_{REV}	EDT_N, P_{REV}	EDT_N, P_{REV}	EDT_N, P_{REV}
B.1 speech	P_{REV}	EDT_N, P_{REV}	EDT_N, P_{REV}	EDT_N, P_{REV}
B.2 cello	—	—	—	—
B.2 speech	P_{REV}	P_{REV}	P_{REV}	P_{REV}
B.3 cello	—	—	—	—
B.3 speech	—	—	—	—
B.4 cello	P_{REV}	P_{REV}	P_{REV}	P_{REV}
B.4 speech	EDT_N, P_{REV}	EDT_N, P_{REV}	P_{REV}	P_{REV}

high positive correlation while no other parameter yields a significant correlation. The comparative effectiveness of P_{REV} for speech can be explained by considering how it is calculated: The auditory model splits a running signal into a direct stream and a reverberant stream, but quasi-stationary ("legato") passages in the music samples provide less opportunity for the algorithm to the reverberant stream than speech samples.

Other issues arising from analysis of a running signal suggest that P_{REV} is more appropriate for estimating a relative change in reverberance, rather than absolute reverberance. For example, P_{REV} in experiments B changes by 15% on average when 1.0 s silence is added to the end of the samples, and by 7% and 14% on average when applying a 1.0 dB and 2.0 dB offset in calibration level, respectively. Note that adding 1.0 s silence changes the L_{Aeq} of the samples, but for this analysis the initial L_{Aeq} values (as in Tables IV–VII) are used in order to explore the influence of this change on P_{REV} when a desired playback level is determined. Recently, the sensitivity of P_{REV} to the SPL was investigated further by Osses Vecchi *et al.* (2017), who showed that the effect of SPL on P_{REV} is dependent on the input spectrum. They suggested further research to investigate the effect of spectral presentation level on reverberance. The outcomes could be used to further improve the model and the robustness of P_{REV}.

However, as long as the same change (e.g., the same calibration offset) is applied to all the samples consistently, the correlation coefficient between P_{REV} and the subjective responses changes by less than 0.02. Furthermore, the fact that P_{REV} is a unitless parameter (so its value is not intuitively interpreted) also suggests the use of P_{REV} for the estimation of a relative change in reverberance. In contrast, T_N and EDT_N are robust to calibration error, and use an intuitive and familiar unit of time in seconds (like T_{20} and EDT). Thus, they appear to be also relevant for estimation of the absolute reverberance.

In Sec. II, the behaviours of T_N, EDT_N, and P_{REV} are compared over a very wide range of SPLs. At very high SPLs, it is reasonable to expect reverberance not to increase with the SPL, in part, because the middle ear reflex provides a 12–14 dB reduction for an intense sound (Howard and Angus, 2013). While the middle ear reflex has a frequency limit and latency (Howard and Angus, 2013), music and speech are sustained enough and have a wide frequency

spectrum for this auditory characteristic to be effective. Based on this assumption, T_N and EDT_N appear to reasonably account for the limited effect of SPL at very high levels. This result is likely attributable to functions in the TVL for the SPL dependency of spectral masking. It should be noted that the TVL and the auditory model used for the calculation of P_{REV} do not include any function for the middle ear reflex.

Overall, the results of this study suggest that further improvements can be made to psychoacoustic reverberance parameters. The calculations of P_{REV} and EDT_N (or T_N) are both based on psychoacoustic modeling, but follow quite different approaches. Hence, the prospects for an improved model that emulates psychoacoustic processes more closely appear to be promising. In future work, it would be interesting to examine how EDT_N and T_N could be extended to model the spatial influences on reverberance using binaural loudness modeling, perhaps augmented by interaural cross correlation. This approach can also be applied to P_{REV} as it processes the two channel outputs of the auditory model (i.e., for left and right ears) separately. Another interesting future work would be to test T_N, EDT_N, and P_{REV} for reverberance of a non-single-exponential sound decay. As shown by Jeong and Joo (2017), the ISO parameters are not accurate in such a condition. The incorporation of human factors in the reverberance estimation, like in T_N, EDT_N, and P_{REV}, should benefit the reverberance estimation for a multi-exponential sound decay.

V. CONCLUSION

The respective performances of T_N, EDT_N, and P_{REV} are compared to each other, as well as to T_{20} and EDT. Results of this study suggest the use of EDT_N for music samples and not for speech (whereas P_{REV} is effective for speech samples). T_N and EDT_N have not been previously tested with speech, and it appears that further development would be needed for them to predict reverberance of speech well. As P_{REV} is derived from a running signal, it is affected by the sample type and the processing of a sample. P_{REV} appears to be excessively sensitive to the SPL. For this reason, P_{REV} is more appropriate for the estimation of a relative change in reverberance for samples of similar SPL, while the RIR-based parameters, T_N and EDT_N, are also good for the estimation of the absolute reverberance. EDT_N is favoured over T_N as a predictor of reverberance. The results of this study identifying the strengths and weaknesses of T_N, EDT_N, and P_{REV} provide a basis for future research into improving these parameters, and provide a general guideline to users of these parameters.

J. Acoust. Soc. Am. **142** (4), October 2017

Lee *et al.* 1839

Barron, M. (**1988**). "Subjective study of British symphony concert halls," Acustica **66**, 1–14.

Breebaart, D. J. (**2001**). "Modelling binaural signal detection," Ph.D. thesis, Eindhoven University of Technology, Eindhoven, The Netherlands.

Breebaart, D. J., van de Par, S., and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.

Cabrera, D., Lee, D., Yadav, M., and Martens, W. L. (**2011**). "Decay envelope manipulation of room impulse responses: Techniques for auralization and sonification," in *Proceedings of Acoustics 2011*, Gold Coast, Australia.

Chalupper, J., and Fastl, H. (**2002**). "Dynamic loudness model (DLM) for normal and hearing-impaired listeners," Acta. Acust. Acust. **88**, 378–386.

Chevret, P., and Parizet, E. (**2007**). "An efficient alternative to the paired comparison method for the subjective evaluation of a large set of sounds," in *Proceedings of the 19th International Congress on Acoustics*, Madrid, Spain.

Denon Professional Test CDs. (**1988**). CD-ROM, Nippon Columbia, Japan.

Farina, A., and Ayalon, R. (**2003**). "Recording concert hall acoustics for posterity," in *Proceedings of the 24th AES Conference*, Banff, Canada.

Fastl, H., and Zwicker, E. (**2007**). *Psychoacoustics: Facts and Models*, 3rd ed. (Springer, Berlin), Chaps. 3, 4, 6, and 8.

Glasberg, B. R., and Moore, B. C. J. (**2002**). "A model of loudness applicable to time-varying sounds," J. Audio. Eng. Soc. **50**, 331–342.

Griesinger, D. (**1997**). "The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces," Acta. Acust. Acust. **83**, 721–731.

Hase, S., Takatsu, A., Sato, S., Sakai, H., and Ando, Y. (**2000**). "Reverberance of an existing hall in relation to both subsequent reverberation time and SPL," J. Sound. Vib. **232**, 149–155.

Howard, D. M., and Angus, J. A. S. (**2013**). *Acoustics and Psychoacoustics*, 4th ed. (Focal, New York), Chap. 2.

ISO 3382-1. (**2009**). "Acoustics—Measurement of room acoustic parameters—Part 1: Performance spaces" (International Organization for Standardization, Geneva, Switzerland).

ITU-R BS. 1284-1. (**2003**). "General methods for the subjective assessment of sound quality" (International Telecommunication Union Radiocommunication Sector, Geneva Switzerland).

Jeong, D., and Joo, H. (**2017**). "Prediction of reverberance in rooms with simulated non-single-exponential sound decays," Appl. Acoust. **125**, 136–146.

Lachenmayr, W. (**2016**). "Loudness-weighting of reverberation using electronic room enhancement," in *Proc. 60th Audio Eng. Soc. Int. Conf.*, Leuven, Belgium.

Lee, D., and Cabrera, D. (**2010**). "Effect of listening level and background noise on the subjective decay rate of room impulse responses: Using time-varying loudness to model reverberance," Appl. Acoust. **71**, 801–811.

Lee, D., Cabrera, D., and Martens, W. L. (**2012a**). "The effect of loudness on the reverberance of music: Reverberance prediction using loudness models," J. Acoust. Soc. Am. **131**, 1194–1205.

Lee, D., Cabrera, D., and Martnes, W. L. (**2012b**). "Accounting for listening level in the prediction of reverberance using early decay time," Acoust. Aust. **40**, 103–110.

Lokki, T., Pätynen, J., Kuusinen, A., and Tervo, S. (**2012**). "Disentangling preference ratings of concert hall acoustics using subjective sensory profiles," J. Acoust. Soc. Am. **132**, 3148–3161.

Mason, R., Brookes, T., and Rumsey, F. (**2004**). "Development of the inter-aural cross-correlation coefficient into a more complete auditory width prediction model," in *Proceedings of the 18th International Congress on Acoustics*, Kyoto, Japan.

Meng, X.-L., Rosenthal, R., and Rubin, D. B. (**1991**). "Comparing correlated correlation coefficients," Psychol. Bull. **111**, 172–175.

Moore, B. C. J., and Glasberg, B. R. (**2007**). "Modeling binaural loudness," J. Acoust. Soc. Am. **121**, 1604–1612.

Morimoto, M., and Asaoka, A. (**2004**). "Multi–dimensional analysis of 'Reverberance'," in *Proceedings of the 18th International Congress on Acoustics*, Kyoto, Japan.

Osses Vecchi, A., Kohlrausch, A., Lachenmayr, W., and Mommertz, E. (**2017**). "Predicting the perceived reverberation in different room acoustic environments using a binaural auditory model," J. Acoust. Soc. Am. **141**, EL381–EL387.

Privitera, G. J. (**2015**). *Statistics for the Behavioural Sciences*, 2nd ed. (Sage, Los Angeles, CA), Chap. 15.

Robinson, D. (**2001**). *Replay gain—A proposed standard*, available at http://wiki.hydrogenaud.io/index.php?title=ReplayGain_specification (Last viewed 7/7/2017).

Rumsey, F. (**2002**). "Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm," J. Audio. Eng. Soc. **50**, 651–666.

Schmitz, A. (**1995**). "Ein neues digitales Kunstkopfmeßsystem" ("A new digital artificial head measuring system"), Acustica **81**, 416–420.

Teret, E., Pastore, M. T., and Braasch, J. (**2017**). "The influence of signal type on perceived reverberance," J. Acoust. Soc. Am. **141**, 1675–1682.

van Dorp Schuitman, J. (**2011**). "Auditory modelling for assessing room acoustics," Ph.D. thesis, Delft University of Technology, Delft, Netherlands.

van Dorp Schuitman, J., de Vries, D., and Lindau, A. (**2013**). "Deriving content-specific measures of room acoustic perception using a binaural, nonlinear auditory model," J. Acoust. Soc. Am. **133**, 1572–1585.