Effect of noise and occupancy on optimal reverberation times for speech intelligibility in classrooms

Murray Hodgson[a] and Eva-Marie Nosal
School of Occupational and Environmental Hygiene and Department of Mechanical Engineering, University of British Columbia, 2206 East Mall, Vancouver, British Columbia V6T 1Z3, Canada

(Received 30 May 2001; revised 15 October 2001; accepted 23 October 2001)

The question of what is the optimal reverberation time for speech intelligibility in an occupied classroom has been studied recently in two different ways, with contradictory results. Experiments have been performed under various conditions of speech-signal to background-noise level difference and reverberation time, finding an optimal reverberation time of zero. Theoretical predictions of appropriate speech-intelligibility metrics, based on diffuse-field theory, found nonzero optimal reverberation times. These two contradictory results are explained by the different ways in which the two methods account for background noise, both of which are unrealistic. To obtain more realistic and accurate predictions, noise sources inside the classroom are considered. A more realistic treatment of noise is incorporated into diffuse-field theory by considering both speech and noise sources and the effects of reverberation on their steady-state levels. The model shows that the optimal reverberation time is zero when the speech source is closer to the listener than the noise source, and nonzero when the noise source is closer than the speech source. Diffuse-field theory is used to determine optimal reverberation times in unoccupied classrooms given optimal values for the occupied classroom. Resulting times can be as high as several seconds in large classrooms; in some cases, optimal values are unachievable, because the occupants contribute too much absorption. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1428264]

PACS numbers: 43.55.Br, 43.55.Hy, 43.55.Ka [JDQ]

I. INTRODUCTION

A major concern regarding the acoustical characteristics of classrooms is speech intelligibility. This is known to be mainly determined by the signal-to-noise level difference (SN)—the difference between the speech-signal and background-noise levels at a receiver—and the amount of reverberation.[1] When speech intelligibility is the concern, the amount of reverberation is best quantified by the early-to-late energy ratio.[2] However, it is more usual to characterize the amount of reverberation in a room by the reverberation time, T, which will be used here.

Speech intelligibility is directly related to signal-to-noise level difference and is inversely related to the reverberation time. However, in rooms the situation is complicated by the fact that reverberation and steady-state levels are interrelated. Increased reverberation, while decreasing the ratio of early-to-late energy ratio to the detriment of speech intelligibility, has the additional effect of increasing steady-state levels by increasing the reverberant sound energy, to the benefit of speech intelligibility.

In this paper, literature on determining optimal reverberation times in classrooms, to optimize speech intelligibility, is reviewed. Fundamental contradictions in the literature are revealed and explained. A new, more physically realistic theoretical approach to predicting the optimal reverberation time in a classroom, considering single and multiple noise sources inside the classroom, is presented and used to corroborate the explanation for the differences in previous results. More realistic optimal reverberation times are derived using the methods developed. How to achieve optimal and satisfactory conditions for speech is discussed. Finally, the optimal reverberation times in the unoccupied classrooms, corresponding to the optimal values in the occupied classroom, are presented.

II. OPTIMAL CLASSROOM REVERBERATION TIMES

A. Literature review

There are two main approaches that have been taken to determine the optimal reverberation time for speech intelligibility in published work—experimental methods and theoretical prediction.

1. Experimental methods

The first approach taken is an experimental one, in which the speech intelligibility of a group of listeners is tested in different acoustical conditions. The conditions that result in the highest speech intelligibility are identified. In one such test by Nabelek and Robinson,[3] modified-rhyme speech tests were recorded in a number of different anechoic and reverberant acoustical environments, and played to test subjects through earphones. A table from the work is reproduced in Table I. The table gives mean word-recognition scores (in percent correct) for monaural and binaural listening at 70 dB speech levels, for various reverberation times and six groups of ten normal-hearing subjects. Note that identification scores decrease as reverberation time increases in all cases; that is, the optimal reverberation time for speech intelligibility is zero. This is as might be expected, since this test procedure did not incorporate noise; it only considered

---

[a] Electronic mail: hodgson@mech.ubc.ca

TABLE I. Mean word-recognition scores in percent found by Nabelek and Robinson (Ref. 3).

Listening mode	Age (years)	$T=0.0$ s	$T=0.4$ s	$T=0.8$ s	$T=1.2$ s
Binaural	10	-	93.9	85.2	87.8
	27	-	98.8	96.7	93.0
	42	-	96.1	91.6	90.2
	54	-	96.1	91.2	88.2
	64	-	89.9	79.8	82.4
	72	-	88.0	80.7	78.0
Monaural	10	99.0	91.6	80.0	82.6
	27	99.7	97.0	92.5	87.7
	42	99.9	91.8	86.9	85.2
	54	99.6	93.3	87.4	83.8
	64	97.2	87.9	70.6	75.6
	72	96.1	83.6	73.1	69.5

the direct effect of reverberation on speech intelligibility, neglecting the effect of signal-to-noise level differences.

Some experimental tests of speech intelligibility did consider signal-to-noise level differences and reverberation together. These tests were performed with the speech and noise generated by one or more loudspeakers at some constant distance from the listener in a room. This approach was taken by Nabelek and Pickett,[4] who found that speech intelligibility decreases with increased reverberation time; that is, they found an optimal reverberation time of zero. Table II is reproduced from their study. It shows the mean perception scores for binaural and monaural hearing through hearing aids. The authors tested five subjects with normal hearing and five subjects with impaired hearing. Results for one normal-hearing subject are given in parentheses. Speech levels for the normal-hearing subjects were presented at 50 dB while those for the hearing impaired were at 60 dB. The results shown are for reverberation times of 0.3 and 0.6 s, and for signal-to-noise level differences of +10 to −15 dB.

Finitzo-Hieber and Tillman[5] did tests similar to those of Nabelek and Pickett, and also found optimal reverberation times of zero. Their results, presented in Table III, give mean monosyllabic word-discrimination scores in percent correct for monaural hearing of normal-hearing children between the ages of 8 and 14 years, for different reverberation times and signal-to-noise differences. Once again, word-discrimination scores decrease as reverberation times increase from zero.

2. Theoretical prediction

The second approach predicts optimal reverberation times from various speech-intelligibility metrics. Values of each metric are determined for various signal and noise levels, and their differences, and for various room characteristics (reverberation time, volume, surface area, absorption, and so forth). The metrics used are considered to be good predictors of speech intelligibility.[1,6] Plomp, Steeneken, and Houtgast[7] predicted nonzero optimal reverberation times by applying method-of-image procedures to predict the modulation transfer function. The audience was assumed to be the main source of noise (with subjects evenly covering the floor at a density of one per square meter, and with each audience member generating a noise level 35 dB below the speech level of the speaker). They found optimal reverberation times increasing from about 0.3 s for a small hall $(10\,\mathrm{m}\times 15\,\mathrm{m}\times 5\,\mathrm{m})$, to about 1.7 s for a large hall $(40\,\mathrm{m}\times 60\,\mathrm{m}\times 20\,\mathrm{m})$!

Bistafa and Bradley[2] used a number of metrics to again predict nonzero values for the optimal reverberation time (that is, the reverberation time that predicted the highest

TABLE II. Mean-word recognition scores in percent found by Nabelek and Pickett (Ref. 4).

		Normal		Impaired	
Listening mode	SN (dB)	$T=0.3$ s	$T=0.6$ s	$T=0.3$ s	$T=0.6$ s
Binaural	Quiet	(98.0)	(95.5)	64.9	57.1
	+10	-	-	60.7	55.2
	+5	-	-	57.4	49.9
	0	-	82.5	49.7	42.4
	−5	76.6	60.7	38.2	31.6
	−10	48.7	38.3	-	-
	−15	26.0	-	-	-
Monaural	Quiet	(95.5)	(95.5)	59.8	54.9
	+10	-	-	57.4	49.6
	+5	-	84.9	54.4	48.9
	0	82.5	72.8	46.7	38.8
	−5	64.0	45.7	33.0	23.3
	−10	25.4	-	-	-

TABLE III. Monosyllabic word-discrimination scores in percent from Finitzo-Hieber and Tillman (Ref. 5).

SN (dB)	$T=0.0$ s	$T=0.4$ s	$T=1.2$ s
∞	94.5	92.5	76.5
12	89.2	82.8	68.8
6	79.7	71.3	54.2
0	60.2	47.7	29.7

value of the metric). Their work assumed diffuse-field theory to be applicable in classrooms. Moreover, they made a number of simplifying assumptions—ratio of classroom volume to surface area equal to 1 m, distance from the speech source to the receiver much greater than the classroom reverberation radius, and negligible air absorption. Table IV is taken from their work; it presents the optimal reverberation which predicts the highest value of the U_{50} speech-intelligibility metric (defined in the following) for rooms of various volumes, V, and signal-to-noise differences, $L_{\mathrm{sf1}}-L_n$. Here, L_{sf1} is the long-term anechoic (free-field) speech level at 1 m directly in front of the talker, and L_n is the total noise level at the listener position. Note that the optimal values are nonzero; an optimal reverberation time of zero is predicted only if $L_{\mathrm{sf1}}-L_n=\infty$, i.e., if there is negligible noise.

The physical explanation for the nonzero optimal reverberation times predicted by Bistafa and Bradley is as follows. In the absence of reverberation, background noise reduces intelligibility. Increased reverberation increases early energy, which can compensate for noise and increase intelligibility. However, too much reverberation decreases the early-to-late energy ratio, decreasing intelligibility.

3. Explaining the contradiction

Why do experimental methods predict an optimal reverberation time of zero while theoretical methods predict nonzero values? Which is correct? The contradictory results for the experimental and theoretical approaches for determining optimal reverberation time require explanation.

It would appear that the difference in the results stems from the different ways in which the methods incorporate noise and, consequently, signal-to-noise level difference. The study by Nebelek and Robinson[3] did not consider the case when noise was present; thus, their results cannot be generalized to conclude that zero reverberation is always optimal. Indeed, Bistafa and Bradley[2] argued that reverberation increases speech levels and, therefore, signal-to-noise level differences, leading to nonzero optimal reverberation times.

By design, the experimental methods employed by Nabelek and Pickett[4] and those by Finitzo-Hieber and Tillman,[5] involved speech and noise generated at positions at the same distance from the listener. More importantly, the signal-to-noise level difference was fixed at the location of the listener's head before each test. In this way, the effect of reverberation on signal-to-noise level differences and, consequently, the possible positive effect of reverberation on speech intelligibility, were eliminated.

TABLE IV. Optimal reverberation times in seconds using U_{50} predicted by Bistafa and Bradley (Ref. 2).

$L_{\mathrm{sf1}}-L_n$ (dB)	$V=100\ \mathrm{m}^3$	$V=300\ \mathrm{m}^3$	$V=500\ \mathrm{m}^3$
10	0.4	0.5	0.6
15	0.3	0.4	0.4
20	0.2	0.3	0.3
25	0.2	0.2	0.2
30	0.1	0.2	0.2

Bistafa and Bradley,[2] on the other hand, assumed total noise levels which did not vary with position or reverberation. The positive effect of reverberation on total speech levels was included by way of a reverberant speech-sound term. However, this method did not consider the analogous negative effect of reverberation on total noise levels. In particular, Bistafa and Bradley used $L_{\mathrm{sf1}}-L_n$ as the input parameter describing signal-to-noise level difference. L_{sf1} is a measure of the speech-source output independent of room effects. In fact, it is analogous to the output power level, L_{ws}; for a point source with directivity index q_s, $L_{\mathrm{sf1}}=L_{\mathrm{ws}}-10\log(q_s)+11$ dB. It was held constant for a given prediction. The effect of the room was taken into account as an increase in speech levels due to reverberation. However, L_n was the total noise level (including the effect of reverberation). Since, for a given prediction, L_{sf1} and $L_{\mathrm{sf1}}-L_n$ were constants, so, effectively, was L_n. Thus, the total noise level did not increase with increasing reverberation; the adverse effect on speech intelligibility, resulting from an increased noise level with increased reverberation time, was not modeled. In principle, the adverse effect of reverberation on noise could be strong enough to cause a decrease in the predicted optimal reverberation times, possibly even to zero.

An alternative way to consider the physical implications of holding L_n constant in the work of Bistafa and Bradley[2] is to assume that the effect of reverberation on noise levels does, in fact, contribute to L_n. Then, as the reverberation time (thus, the reverberant-field contribution to L_n) increases, the power of the noise source effectively decreases, to keep L_n constant. From this perspective, it is likely the fact that the noise-source output effectively decreases with increasing reverberation time, which leads to predictions of nonzero optimal reverberation time. In reality, however, the inherent sound-level output of the noise source does not change. This begs the question of whether this method would still predict nonzero optimal reverberation times if the noise-source output were kept constant?

The method-of-images approach of Plomp, Steeneken, and Houtgast[7] allowed arbitrary (rectangular) room geometries and absorption coefficients, thus considerably reducing the limitations associated with diffuse-field theory, by predicting the spatial propagation characteristics of the rooms more realistically. In their work, noise was incorporated by considering the audience as a collection of individual noise sources.

B. Theoretical prediction of optimal reverberation time

1. Basic equations

To address the issue of the differences in the predicted optimal reverberation times, we focus on noise sources inside the classroom (projectors, ventilation outlets, and occupants), and consider the U_{50} speech-intelligibility metric that has been shown to be best suited for the evaluation of speech intelligibility in classrooms.[1] U_{50} is among the metrics that make use of the acoustical energy-ratio concept. This concept divides received acoustical energy into useful and detrimental parts. The useful part consists of the direct energy from the speaker, E_d, and the early arriving, reflected energy from the speaker, E_e. In U_{50}, reflected energy that arrives at the listener within 0.05 s of the signal is classified as early (useful) reflected energy. The remaining reflected, or late-arriving, energy, E_l, is considered detrimental. In addition to late-arriving reflected energy, noise energy, E_n, is included as detrimental. In summary, the useful-to-detrimental energy ratio Q is defined as

$$Q=\frac{E_d+E_e}{E_l+E_n}. \tag{1}$$

U_{50} in decibels is given by

$$U_{50}=10\log(Q). \tag{2}$$

To determine Q, let us follow previous work[2] by considering each of the four components of energy defined previously. Assuming point sources, and applying diffuse-field theory which assumes an exponential sound decay and a spatially invariant reverberant field, it can be shown that the direct and total-reflected energy densities ($E_r=E_c+E_l$) in w s/m^3 are, respectively, given by[8]

$$E_d=\frac{q_s W}{4\pi c r_s^2}, \tag{3}$$

and

$$E_r=\frac{q_s W}{4\pi c r_h^2}, \tag{4}$$

where q_s is the directivity index of the source, W is the output sound power of the point acoustic source (assumed, for now, to be a speaker) in w, c is the speed of sound in m/s, r_s is the distance from the source to the listener in m, and r_h is the reverberation radius in m. The reverberation radius is that distance at which the total reflected sound energy equals the direct sound energy ($E_d=E_r$). To find r_h, consider Eyring's reverberation formula:

$$T=\frac{24V\ln(10)}{c[4mV-S\ln(1-\alpha)]}, \tag{5}$$

where T is the reverberation time in s, V is the room volume in m^3, S is the room surface area in m^2, α is the average surface absorption coefficient, and m is the air-absorption exponent in Np/m. After appropriate manipulation we obtain

$$1-\alpha=\exp[(4V/S)(m-k/c)], \tag{6}$$

with $k=\ln(10^6)/T$.

Using the formula for reflected energy density given by Kuttruff:[8]

$$E_r=-\frac{4W(1-\alpha)}{S\ln(1-\alpha)}, \tag{7}$$

and setting $E_d=E_r$ with $r_s=r_h$ in the expression for E_d, then solving for r_h^2 we obtain:

$$r_h^2=\frac{qV(k-mc)\exp[(4V/S)(k/c-m)]}{4\pi c}. \tag{8}$$

It can further be shown[1] that

$$E_e=E_r(1-e^{-k/20}), \tag{9}$$

from which it follows that

$$E_l=E_r e^{-k/20}. \tag{10}$$

Now consider L_{sf1}, the long-term anechoic speech level at 1 m directly in front of the speaker (analogous to the speaker output-power level). If L_d denotes the direct sound pressure level, and E_s denotes the sound energy density in w s/m^3 at 1 m, we have

$$L_d-L_{\mathrm{sf1}}=10\log\left(\frac{E_d}{E_s}\right)=10\log\left(\frac{1}{r_s^2}\right) \tag{11}$$

from which it follows that

$$\frac{\rho c^2 E_d}{p_0^2}=\frac{10^{L_{\mathrm{sf1}}/10}}{r_s^2} \tag{12}$$

with $p_0=2\times10^{-5}$ Pa. Similarly:

$$\frac{\rho c^2 E_r}{p_0^2}=\frac{10^{L_{\mathrm{sf1}}/10}}{r_{h_s}^2}. \tag{13}$$

2. Improved treatment of noise

In this improved analysis, noise due to a source inside the classroom is treated as generated by a point source at distance r_n from the listener. First, we note from Eq. (8) that

$$\frac{r_{h_s}^2}{q_s}=\frac{r_{h_n}^2}{q_n}, \tag{14}$$

where the subscripts s and n refer to the speech and noise sources, respectively. Then, following the same steps as for the speaker, we find:

$$\frac{\rho c^2 E_n}{p_0^2}=\frac{\rho c^2}{p_0^2}(E_{\mathrm{nd}}+E_{\mathrm{nr}})=10^{L_{\mathrm{nf1}}/10}\left(\frac{1}{r_n^2}+\frac{1}{r_{h_n}^2}\right), \tag{15}$$

where E_{nd} is the direct-energy density from the noise source, E_{nr} is the reflected-energy density from the noise source, and L_{nf1} is the long-term anechoic noise level at one meter directly in front of the noise source (analogous to the noise-source output-power level). Combining the above-given results and simplifying yields

$$U_{50}=10\log\left(\frac{(r_{h_s}^2/r_s^2)+1-e^{-k/20}}{e^{-k/20}+10^{(L_{\mathrm{nf1}}-L_{\mathrm{sf1}})/10}\left(\frac{r_{h_s}^2}{r_n^2}+\frac{q_s}{q_n}\right)}\right). \tag{16}$$

This equation allows for the calculation of U_{50} as a function of reverberation time, speaker-to-listener distance and speech-source output level, noise-source-to-listener distance and noise-source output level, and speech- and noise-source directivity indices, for a given room (in particular, for a given volume, surface area and air-absorption exponent). Thus, given any room and its characteristics, and given the distances of the speech and noise sources from the listener, and their directivities, we can predict the optimal reverberation time for given speech- and noise-source output levels, by finding the value of T that maximizes U_{50}. Note, however, that Eq. (16) is still based on diffuse-field theory; in particular, it assumes a spatially invariant reverberant field.

It is important to emphasize here that, in Eq. (16), signal-to-noise level difference is defined in terms of values inherent to the speech- and noise-source outputs, independent of their acoustical environments, and not, as is common, in terms of the values at a receiver position which are, of course, strongly dependent on the acoustical environment. Source-related values will be indicated by SNS, received values by SNR. Output-power levels of typical speakers can be determined from published data;[9] total A-weighted values vary with vocal effort from about 60 to 75 dB. Output-power levels of UBC-classroom projectors and ventilation outputs have been measured.[10] Typical total A-weighted values are in the range 40–60 dB. These data suggest that SNS would be expected to range from about 0 to 40 dB in typical classrooms.

3. U_{50} prediction

U_{50} predictions were done using Eq. (16) for six classrooms with volumes $V=50$, 100, 300, 500, 1000, and 4000 m^3, the range found at the University of British Columbia (UBC). Corresponding surface areas were determined from $S=5.36\ V^{0.7104}$, the equation describing the best-fit regression between V and S for 279 UBC classrooms involved in another study.[11] Various distances from the speech source and the noise source to the receiver were used, as detailed in the following. Predictions were made for SNS=0, 10, 20, 30, and 40 dB (again, $\text{SNS}=L_{\text{sf1}}-L_{\text{nf1}}$; the actual noise- and speech-source output levels do not matter). The values of the other prediction parameters were as follows: $q_s=2$, $q_n=1$, $m=0.0012$ Np/m, $c=344$ m/s. Note that noise sources were assumed to be omni-directional, an assumption that may not be accurate. Ventilation ducts and projectors would be expected to be directional at higher frequencies. However, their orientations in the classroom, and the positions of receivers with respect to them, are highly variable. Thus, an assumption of omni-directionality is a reasonable one to make from a practical point of view.

Two interesting general results were found when U_{50} was maximized for various rooms and for various speech and noise levels and listener distances:

(1) In any case for which the speech source was at a distance equal to or less than the distance from the noise source to the listener, an optimal reverberation time of zero was predicted.

(2) When the noise source was closer to the listener than the speaker, nonzero optimal reverberation times were, in general, predicted. The optimal reverberation time increased with classroom volume. It decreased with increased signal-to-noise level difference, tending to zero as the level difference tends to infinity. For a given r_s, the optimal reverberation time increased with decreasing r_n.

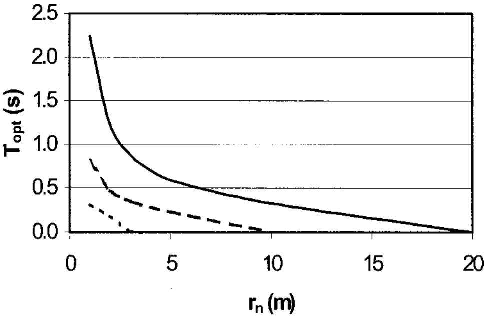

FIG. 1. Predicted variation of the optimal reverberation time T_{opt} with the distance from the noise source to the receiver, r_n, for classrooms with three volumes: (—) 4000 m^3; (---) 500 m^3; (···) 50 m^3, and SNS=0 dB. The speaker is at the front, the receiver at the back, of the classroom.

Figures 1 and 2 illustrate the results. Figure 1 represents a "worst-case" scenario with large r_s and small r_n. Here r_s was chosen to represent the largest likely distance from the speech source to the receiver—that for a speaker at the front of the classroom and the receiver at the back. Values of r_s were calculated as a function of volume using $r_s=0.7145\ V^{0.417}$, the equation describing the best-fit regression between V and the distance to the center, rear of the classroom (seating area F) for 279 UBC classrooms involved in another study;[11] the values used are shown in Table V. The noise-source-to-receiver distance was $r_n=1$ m corresponding, for example, to a receiver seated close to a noisy ventilation outlet or projector. Shown in Fig. 1 are the optimal reverberation times in the classrooms with $V=50$, 500, and 4000 m^3, for the case of SNS=0 dB. Figure 2 shows the same results for a receiver seated at distances one half those used in Fig. 1. Table V summarizes the worst-case predictions of optimal reverberation time. The optimal reverberation time increases with increasing volume from about 0.3 s to several seconds with low signal-to-noise level difference, and from 0.1 to 0.3 s with near optimal signal-to-noise level difference. In fact, the results are similar to, but slightly lower than, those of Bistafa and Bradley[2]—see Table IV.

These results can be explained physically in an approximate way as follows. A given increase in reverberation, while inherently tending to decrease speech intelligibility,

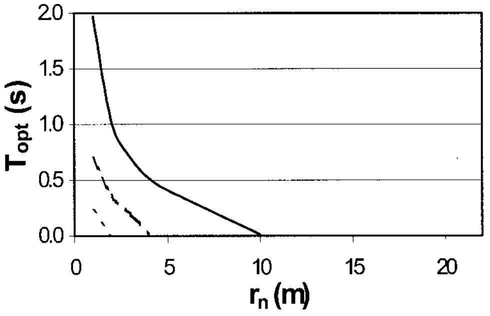

FIG. 2. Predicted variation of the optimal reverberation time T_{opt} with the distance from the noise source to the receiver, r_n, for classrooms with three volumes: (—) 4000 m^3; (---) 500 m^3; (···) 50 m^3, and SNS=0 dB. The speaker is at the front, the receiver in the middle, of the classroom.

TABLE V. Optimum reverberation times (T_{opt} in s—first line), received signal-to noise level differences (SNR in dB—second line) and U_{50}'s in dB (third line), determined using Eq. (16). Also shown are the optimum average surface-absorption coefficients (α_{opt}—fourth line) calculated from T_{opt} using diffuse-field theory.

SNS (dB)	Quantity	$V=50$ m^3 $r_s=3.7$ m	$V=100$ m^3 $r_s=4.9$ m	$V=300$ m^3 $r_s=7.7$ m	$V=500$ m^3 $r_s=9.5$ m	$V=1000$ m^3 $r_s=12.7$ m	$V=4000$ m^3 $r_s=22.7$ m
	T_{opt} (s)	0.08	0.08	0.10	0.11	0.12	0.17
40	SNR (dB)	34.23	31.35	27.35	25.42	22.60	17.62
	U_{50} (dB)	30.73	28.13	23.91	21.91	19.20	13.93
	α_{opt}	0.69	0.76	0.79	0.81	0.84	0.86
	T_{opt} (s)	0.10	0.14	0.14	0.16	0.19	0.29
30	SNR (dB)	25.03	22.65	18.92	17.22	14.83	10.47
	U_{50} (dB)	21.61	19.22	15.23	13.36	10.82	5.94
	α_{opt}	0.61	0.64	0.67	0.68	0.69	0.68
	T_{opt} (s)	0.14	0.17	0.23	0.27	0.35	0.61
20	SNR (dB)	16.18	14.49	11.43	10.02	8.28	4.97
	U_{50} (dB)	12.69	10.69	7.22	5.56	3.33	−1.08
	α_{opt}	0.48	0.49	0.49	0.49	0.47	0.41
	T_{opt} (s)	0.23	0.29	0.43	0.53	0.72	1.42
10	SNR (dB)	7.54	6.38	4.25	3.25	1.92	−0.54
	U_{50} (dB)	3.85	2.30	−0.56	−2.01	−4.06	−8.42
	α_{opt}	0.33	0.32	0.30	0.29	0.26	0.20
	T_{opt} (s)	0.34	0.43	0.72	0.91	1.27	2.85
0	SNR (dB)	−1.69	−2.58	−3.95	−4.69	−5.73	−7.39
	U_{50} (dB)	−5.43	−6.69	−9.22	−10.59	−12.62	−17.09
	α_{opt}	0.24	0.23	0.19	0.18	0.15	0.10

also results in a constant position-invariant increase in steady-state reverberant speech and noise levels which add to direct levels to give the total speech and noise levels. For both the speech and the noise, the magnitude of the total-level increases depends on the relative contributions of the direct and reverberant levels. At low source/receiver distance, the direct sound dominates, and the increase is small; at large distance the reverberant sound dominates and the increase is large. Thus, the constant increase in reverberant level results in an increase in the total level which increases with source/receiver distance. The relative amounts by which the speech and noise levels increase with increased reverberation determine whether the resulting speech-to-noise level difference increases (tending to increase intelligibility), decreases (tending to decrease intelligibility), or remains the same. This depends on the relative distances of the speech and noise sources from the receiver. If the distances are the same, both the speech and noise levels vary with reverberation in the same way, and the level difference does not change; the only effect on intelligibility is a decrease due to increased reverberation. If, on the other hand, the noise source is farther from the receiver than the speech source, then noise levels increase more with reverberation than do speech levels; resulting level differences decrease, tending to decrease intelligibility. This reinforces the effect of reverberation on intelligibility, and leads to an optimal reverberation time of zero. Finally, if the speech source is farther from the receiver than the noise source, then speech levels increase more with reverberation than do noise levels, and the level difference increases with reverberation, tending to increase intelligibility. This trend tends to counteract the decrease in intelligibility due to temporal reverberation, leading to the highest intelligibility occurring at a nonzero optimal reverberation time.

It is of interest to consider the results in terms of the more commonly used received signal-to-noise level difference, SNR. Values corresponding to T_{opt} are shown in Table V. It is generally considered that SNR>15 dB is required for excellent speech intelligibility.[2] Table V shows that SNSs greater than about 20, 28, and 38 dB, respectively, are required to achieve this in small, medium, and large classrooms.

It is also of interest to consider the results in terms more directly relevant to classroom design—for example, the amount of sound absorption required to achieve the optimal reverberation times. Table V shows the optimal average surface-absorption coefficients, α_{opt}, corresponding to the T_{opt} values, calculated using diffuse-field theory. In quiet classrooms, requiring low reverberation times and, in general, high absorption, the optimal coefficient increases with volume from about 0.7 to 0.9. In moderately noisy classrooms, it varies between about 0.4 and 0.5. In noisy classrooms, requiring higher reverberation and, generally, low absorption, α_{opt} decreases with volume from about 0.2 to 0.1 s. The classroom sound-absorptive features that can be used to achieve these various values of α_{opt} can be determined from absorption-coefficient data published elsewhere.[12,13]

4. Reverberation time for satisfactory speech intelligibility

It should be noted that, although the reverberation times recommended previously do not provide optimal speech intelligibility for listeners closer to the speaker than to the

TABLE VI. Ranges of sufficient reverberation time in seconds for satisfactory speech intelligibility, for the "worst case" of large r_s and $r_n=1$ m. "Na" indicates that satisfactory speech intelligibility cannot be achieved.

SNS (dB)	$V=50$ m^3 $r_s=3.7$ m	$V=100$ m^3 $r_s=4.9$ m	$V=300$ m^3 $r_s=7.7$ m	$V=500$ m^3 $r_s=9.5$ m	$V=1000$ m^3 $r_s=12.7$ m	$V=4000$ m^3 $r_s=22.7$ m
40	0–0.9	0–0.9	0–0.9	0–0.9	0–0.9	0–0.9
30	0–0.9	0–0.9	0–0.9	0–0.9	0–0.9	0–0.9
20	0–0.8	0–0.8	0–0.8	0.1–0.8	0.2–0.8	Na
10	0.1–0.6	0.2–0.5	Na	Na	Na	Na
0	Na	Na	Na	Na	Na	Na

noise source (we have found optimal values of zero in this case), U_{50} tends to be high for these listeners. In Table V, U_{50} values corresponding to the optimal reverberation times are shown; they are as higher than 30. Optimal reverberation times may not be necessary for satisfactory speech intelligibility. For example, it has been suggested that $U_{50}\geqslant 1.0$ is sufficient for satisfactory speech intelligibility.[1] Following this suggestion, Table VI gives the ranges of reverberation times which result in satisfactory speech intelligibility for the same worst-case configurations considered previously. $U_{50}\geqslant 1.0$ can generally be achieved with SNS greater than about 15 dB, and T in the range 0 to 0.9 s; it generally cannot be achieved if SNS<0 dB. "Na" entries indicate conditions for which $U_{50}=1.0$ cannot be obtained, regardless of reverberation time. Table V indicates the combinations of classroom volume and source-related signal-to-noise level difference, SNS, which correspond to $U_{50}\geqslant 1.0$. In classrooms with low, medium, and high volume SNS must exceed about 5, 15, and 25 dB, respectively, to achieve satisfactory speech intelligibility.

5. Extension to multiple noise sources

Classrooms usually contain only one main speech source, but they may contain many sources of noise (e.g., ventilation outlets, projectors, occupants). The abovementioned work assumed only one noise source. However, we can account for multiple noise sources (say, m of them, where n_i refers to the ith noise source) in the present analysis by simply adding the energy contributions E_{ni} ($i=1,2,...,m$) of all sources to give the total-noise energy density:

$$E_n=\sum_{i=1}^{m} E_{ni}=\sum_{i=1}^{m} 10^{L_{\mathrm{nf1}i}/10}\left(\frac{1}{r_{ni}^2}+\frac{1}{r_{h_{ni}}^2}\right), \tag{17}$$

where $L_{\mathrm{nf1}i}$ is the long-term anechoic noise level at 1 m directly in front of the ith noise source, and r_{ni} is the distance of the ith noise source to the listener. This gives

$$Q=\frac{(r_{h_s}^2/r_s^2)+1-e^{-k/20}}{e^{-k/20}+\sum_{i=1}^{m} 10^{(L_{\mathrm{nf1}i}-L_{\mathrm{sf1}})/10}\left(\frac{r_{h_s}^2}{r_{ni}^2}+\frac{q_s}{q_{ni}}\right)}. \tag{18}$$

When this method is applied to multiple noise sources, those further from the listener than the signal have negligible effect in the prediction of optimal reverberation time. For example, when a noise source (regardless of output level) is added at 30 m from the listener in the example in the previous section (where the signal was 20 m from the listener and the first noise source was 1 m from the listener), optimal reverberation-time predictions remain exactly the same. This result suggests that, for multiple noise sources, only those closer to the listener than the speech source need to be considered. Also, as might be expected, adding noise sources that are closer to the listener than the signal results in an increase in the predicted optimal reverberation time.

6. Further explanation of the contradiction in the literature

Different treatments of noise were suggested as the reason why empirical methods find an optimal reverberation time of zero while theoretical methods predict nonzero values. We can use Eq. (16) to corroborate this explanation. First, as noted previously, experimental methods effectively ignore the positive effect of reverberation on signal-to-noise level differences. This is equivalent to replacing the $(r_{h_s}^2/r_n^2+1)$ term in the denominator of Eq. (16) with one, resulting in optimal reverberation times of zero. If the signal-to-noise level difference were allowed to vary with changing reverberation, the design of experimental methods (with noise and speech generated at the same distances from the listener) would correspond to $r_n=r_s$. In this case, Eq. (16) predicts an optimal reverberation time of zero.

In the treatment of noise by Bistafa and Bradley,[2] the r_n variable was eliminated, because there was no source of noise. Also, the effect of reverberation on noise was included in L_{nf1}, which was the total noise level. This is equivalent to replacing the $(r_{h_s}^2/r_n^2+1)$ term in the denominator of Eq. (16) with $r_{h_s}^2$. To apply our model in this case, we treat the uniform and stationary noise as originating from a noise source close to the listener—say $r_n=1$ m. Then $r_n\ll r_s$, and Eq. (16) predicts nonzero optimal reverberation times (see Table V).

In the work of Plomp, Steeneken, and Houtgast,[7] a listener had several point noise sources nearby. In particular, there was a noise source that was closer to the listener than the speaker. From our discussion of multiple noise sources, we see that our method predicts nonzero optimal reverberation times in this situation. This is consistent with the findings of Plomp, Steeneken, and Houtgast.

III. OPTIMAL REVERBERATION TIMES IN UNOCCUPIED CLASSROOMS

The optimal classroom reverberation times discussed previously (e.g., as presented in Table V) are those experienced by the occupants of the classroom when in use—that

TABLE VII. Predictions of optimal reverberation time in seconds in unoccupied classrooms from the values when occupied from Table V. N is the number of classroom occupants.

SNS (dB)	$V=50\ \mathrm{m}^3$ $N=12$	$V=100\ \mathrm{m}^3$ $N=25$	$V=300\ \mathrm{m}^3$ $N=70$	$V=500\ \mathrm{m}^3$ $N=115$	$V=1000\ \mathrm{m}^3$ $N=230$	$V=4000\ \mathrm{m}^3$ $N=400$
40	0.1	0.1	0.1	0.1	0.1	0.2
30	0.1	0.1	0.2	0.2	0.2	0.3
20	0.2	0.2	0.3	0.4	0.6	0.9
10	0.3	0.5	0.9	1.4	4.3	5.0
0	0.6	0.9	4.7	-	-	-

is, in the occupied classroom. It would be convenient, however, to be able to specify optimal reverberation times for the unoccupied classroom, since unoccupied values are those to which designers can more readily design. This can be accomplished by calculating the reverberation times in the unoccupied classrooms from those that are optimal in the occupied classrooms, given the number of occupants and their absorption characteristics. This can be done using diffuse-field theory, according to which the occupied reverberation time T_o in seconds is given by

$$T_o=\frac{0.161V}{A_u+NA_p}, \tag{19}$$

in which V is classroom volume in m^3, A_u is the unoccupied room absorption in m^2, N is the number of occupants, and A_p is the absorption per occupant (the 1 kHz value of 0.81 m^2 is used here).[13]

Now, according to diffuse-field theory, for the unoccupied classroom, $T_u=0.161V/A_u$; thus, $A_u=0.161V/T_u$. Substituting this into Eq. (19) and rearranging gives

$$T_u=\frac{1}{\frac{1}{T_o}-\frac{A_pN}{0.161V}}. \tag{20}$$

Note that, according to this expression, classrooms of different volumes with the same occupied T_o's and the same numbers of occupants per unit volume, have the same unoccupied T_u's (and vice versa).

Consider classrooms with volumes of 50, 100, 300, 500, 1000, and 4000 m^3, as discussed previously. Assume that these classrooms contain 12, 25, 70, 115, 230, and 400 occupants, respectively (these were again determined from typical data for 279 UBC classrooms considered in another study[11] on the assumption of 70% occupancy, typical of UBC classrooms). Table VII shows, for various source-related signal-to-noise level differences, the optimal unoccupied T_u 's associated with the optimal occupied T_o's in Table V. In the case of the low optimal T_o's associated with high signal-to-noise level difference, the optimal T_u's are almost the same. However, in the case of the higher optimal T_o's associated with low signal-to-noise level difference and large volume, optimal T_u's can be as high as several seconds. In several of these cases the optimal T_u was negative (in Table VII, this is indicated by a dash). The explanation of this nonphysical result is that the amount of absorption provided by the classroom occupants exceeded that required to obtain the optimal occupied reverberation time; that is, the optimal occupied reverberation time cannot be achieved if the classroom is 70% occupied.

IV. CONCLUSION

Our work indicates that the question of optimal reverberation time is ultimately reduced to the question of how to incorporate noise in a physically realistic manner. Here we have considered noise sources located in the classrooms. To treat noise as it is treated in the experimental methods—where its level is adjusted with changing reverberation to keep the signal-to-noise level difference at the listener constant—is not realistic. In a real classroom, it is the inherent output-power level of the noise source and of the speaker that is constant, regardless of reverberation, not the levels at the listener position. Nor is it realistic to treat noise as generated at a single point somewhere within the classroom at a distance from the listener equal to the distance between the speaker and the listener. Sources of ventilation and student-activity noise are very easily, and most likely, located much closer to the listener than is the speaker. Consequently, the way in which noise is incorporated in the theoretical methods is, in this respect, more realistic than in experimental methods. In using theoretical methods, however, we must still be careful to incorporate the effect of reverberation on noise. Bistafa and Bradley[2] did not do this, while Plomp, Steeneken, and Houtgast[7] did.

A physically realistic treatment of noise incorporates both the nearby noise source and the effect of reverberation on noise—for example, by setting $r_n<r_s$ in Eq. (16). The results of such an analysis are given in Table V. We see that when noise is incorporated in a more physically realistic manner, nonzero reverberation times, in the range of 0.1 s to several seconds, are found to be optimal.

There are numerous reasons why a reverberation time of zero may not, in practice, be desirable. First, the cost of reducing reverberation times to very low values may be prohibitively expensive and impractical. Second, it is unnatural to be in an environment in which there is very little reverberation. A listener in such an environment may feel uncomfortable, while speakers may have difficulty monitoring their voices (since no energy is returned). Also, since speech is directional (directed predominantly in front of the speaker), listeners in an anechoic situation who are not positioned directly in front of the speaker will receive relatively little direct speech signal. Further, since the rate of spatial decrease of direct energy is inversely proportional to the square of the distance between the listener and the speaker, listeners far from the speaker (near the back of a large classroom, for example) will also receive little direct speech signal. These listeners rely on reflected sound energy to hear what is being said.

For all of the above-mentioned reasons—in particular, because physically realistic incorporation of noise leads to predictions of nonzero optimal reverberation time—it is safe to conclude that one should aim for nonzero reverberation times in occupied classrooms. Reverberation times varying from 0 to 1 s with increasing noise level and classroom size appear to be appropriate; these correspond to optimal values in unoccupied classrooms which are as high as several seconds. However, optimal values may, in fact, be impossible to achieve in well-occupied classrooms, since the absorption provided by the occupants may exceed that required for optimal reverberation times.

The above-presented work is by no means conclusive and points to other areas that should be explored. The use of other metrics to predict optimal reverberation time is one natural extension of the work done here with U_{50}. Although such work has been done,[2] it would be beneficial to seek results when noise is incorporated in a more physically realistic manner and which eliminates the assumption of diffuse-field theory. To do so, and to improve the current work, it is necessary to accurately model noise in classrooms (or, if this has already been done, to apply these models to the predictions). Another interesting approach would be to realistically incorporate noise into the experimental methods. For obvious reasons, this may be quite difficult (or even impractical) and, again, demands an accurate model of noise in classrooms.

It is also important to note that increased reverberation times and background noise may have more adverse effects on children, older listeners, and the hearing impaired than on normal-hearing, adult listeners.[3,5] Consequently, the above-presented results may not apply to these listeners. Also of importance would be to find U_{50} levels which correspond to acceptable speech intelligibility for various groups of special listeners (children, elderly, hearing impaired), and to use these to suggest ideal and acceptable reverberation levels in classrooms designed to teach these people.

Finally, further work is required to revise the prediction models to reduce their reliance on diffuse-field theory. In particular, it would be of interest to incorporate more realistic decreases of steady-state levels with source–receiver distance.[12,13]

[1] J. S. Bradley, ‘‘Speech intelligibility studies in classrooms,’’ J. Acoust. Soc. Am. **80**, 846–854 (1986).

[2] S. R. Bistafa and J. S. Bradley, ‘‘Reverberation time and maximum background-noise level for classrooms from a comparative study of speech intelligibility metrics,’’ J. Acoust. Soc. Am. **107**, 861–875 (2000).

[3] A. K. Nabelek and P. K. Robinson, ‘‘Monaural and binaural speech perception in reverberation for listeners of various ages,’’ J. Acoust. Soc. Am. **71**, 1242–1248 (1982).

[4] A. K. Nabelek and J. M. Pickett, ‘‘Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners,’’ J. Speech Hear. Res. **7**, 724–739 (1974).

[5] T. Finitzo-Hieber and T. W. Tillman, ‘‘Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children,’’ J. Speech Hear. Res. **21**, 440–458 (1983).

[6] J. S. Bradley, ‘‘Predictors of speech intelligibility in rooms,’’ J. Acoust. Soc. Am. **80**, 837–845 (1986).

[7] T. Houtgast, H. J. M. Steeneken, and R. Plomp, ‘‘Predicting speech intelligibility in rooms from the modulation transfer function. II. Mirror image computer model applied to rectangular rooms,’’ Acustica **46**, 73–81 (1980).

[8] H. Kuttruff, *Room Acoustics*, 3rd ed. (Elsevier Applied Science, London, 1991), pp. 114–123.

[9] ‘‘American National Standard Methods for Calculation of the Speech Intelligibility Index,’’ ANSI S3.5-1997.

[10] R. Wang, ‘‘Sound-intensity measurements of the sound-power outputs of classroom projectors and ventilation outlets,’’ Technical report, University of British Columbia, 2001.

[11] M. R. Hodgson, ‘‘Rating, ranking and explaining the acoustical quality of university classrooms,’’ J. Acoust. Soc. Am. (submitted).

[12] M. R. Hodgson, ‘‘Empirical prediction of speech levels and reverberation in classrooms,’’ J. Build. Acoust. **8**, 1–14 (2001).

[13] M. Hodgson, ‘‘Experimental investigation of the acoustical characteristics of university classrooms,’’ J. Acoust. Soc. Am. **106**, 1810–1819 (1999).