Optimum acoustical conditions for speech in classrooms
Bistafa, S. R.; Bradley, J. S.

Publisher's version / Version de l'éditeur:

Noise and Vibration Worldwide, 31, October 9, pp. 12-17, 2000-10-01

Optimum acoustical conditions for speech in classrooms

NRCC-48327

Bistafa, S.R.; Bradley, J.S.

October 200

A version of this document is published in / Une version de ce document se trouve dans:
Vibration Worldwide, 31, (9), October, pp. 12-17, October 01, 2000

Optimum acoustical conditions for speech intelligibility in classrooms

Sylvio R. Bistafa
Department of Mechanical Engineering, Polytechnic School, University of Sao Paulo, Sao Paulo, 05508-900, SP, Brazil.

John S. Bradley
Institute for Research in Construction - Acoustics, National Research Council, Ottawa, K1A 0R6, ON, Canada.

Reverberation time and the maximum background-noise level are the usual basis of classroom acoustic standards and regulations. Comparing the results provided by three speech intelligibility metrics, under the assumption of a diffuse sound field with ideal exponential decays, led to a range of estimates of optimum conditions for optimum speech intelligibility. A reverberation time of 0.4 s and a maximum background noise level 25 dB below the voice level measured at 1 m in front of the talker were indicated to lead to optimum speech intelligibility by all three speech metrics and are therefore suggested to ensure the best speech intelligibility conditions in classrooms.

Introduction

The ultimate acoustical goal in a classroom is adequate speech intelligibility. Sound reflections and background noise control speech intelligibility in rooms. Physical measures of speech intelligibility here referred to as speech intelligibility metrics, have been proposed to measure the combined effects of room reflections and background noise on speech intelligibility. Despite the fact that most speech intelligibility metrics can be measured in real rooms, the classroom acoustical performance is seldom specified in terms of these quantities. Reverberation time and the maximum background-noise level are the usual basis of classroom acoustic standards and regulations.

The objective of the present work is to obtain the reverberation time and the maximum background-noise level for classrooms for best speech intelligibility conditions. These will be derived by comparing the results provided by three speech intelligibility metrics, under the assumption of a diffuse sound field with ideal exponential decays.

1. Speech intelligibility metrics

There are several speech metrics for the evaluation of speech intelligibility conditions in rooms [1]. The three most used speech metrics which take into account the effects of room reflections and background noise on speech intelligibility are the useful-to-detrimental sound ratio U_{50} [2], the Speech Transmission Index STI [3] and formulas for predicting the articulation loss of consonants AL_{cons} [4].

The speech metrics values can be obtained from measurements of room impulse responses. It is possible however, to derive analytical expressions for the speech metrics under the assumption of a diffuse sound field with ideal exponential decays. Although a diffuse sound field is only approximated in real rooms, it provides the basis for the theoretical determination of parameters important for the evaluation of speech intelligibility conditions. Thus, under the diffuse sound field assumption, the acoustics of the room can be described by the reverberation time T and the signal-to-noise ratio.

Classrooms are relatively small rooms for speech with volumes V up to 500 m^3 or so. They are in general rectangular-shaped rooms with volume-to-surface-area ratios $V/S \approx 1$m. The sound source is the human voice without amplification, which can be acoustically specified in terms of the long-term averaged speech level at 1m L_{splm}, with an average directivity factor at mid-frequencies $q = 2$ straight ahead of the talker.

Under the assumption of a diffuse sound field in the room, an expression for U_{50} can be found in the form

$$U_{50} = 10\lg\left[\frac{1-e^{-0.69/T}}{e^{-0.69/T}+10^{(L_n-L_r)/10}}\right]. \quad (1)$$

The STI is calculated using a procedure [3], based on the Modulation Transfer Function m(F). Under the assumption of a diffuse sound field in the room, m(F) is given by

$$m(F) = \left[1+\left(\frac{2\pi}{13.82}F\cdot T\right)^2\right]^{-1/2}\cdot\left[1+10^{(L_n-L_r)/10}\right]^{-1}, \quad (2)$$

where F is the speech modulation frequency.

The articulation loss of consonants can be predicted by the so-called architectural form of the Peutz's equation [4], which is given by

$$\%AL_{cons} = 9T(1.071\cdot T^{-0.0285})^{25+(L_n-L_r)}, \quad (3)$$

where $\%AL_{cons}$ is the articulation loss of consonants as a percentage. Equation (3) is valid for $L_n - L_r \geq -25$ dB.

Equations (1) – (3) are written for distances greater than the limiting distance, for which the contribution of the direct field is of no significance. The limiting distance can be defined as the distance for which the direct-to-reflected sound energy density is equal to -10dB. For a 300 m^3 classroom in which the reverberation time is 0.5 s, the limiting distance is approximately 7.3 m. This is a worst case scenario since at distances less than the limiting distance there is an increasing contribution from the direct field as the receiver approaches the speaker, which tends to improve the speech intelligibility conditions.

In Eqs. (1) and (2), $10^{(L_n-L_r)/10}$ is the noise-to-reflected-speech ratio, which for $V/S = 1$m can be written as

$$10^{(L_n-L_r)/10} = 0.0032\cdot q\cdot V\cdot T^{-1}\cdot e^{0.16/T}\cdot 10^{(L_n-L_{splm})/10} \quad (4)$$

It should be noticed that Eqs. (1) – (4) have all been written in terms of the reverberation time T, which, under the assumption of a diffuse sound field with ideal exponential decays, is the only quantity that describes the room acoustical conditions.

Some earlier developments of speech intelligibility metrics were based on unfiltered broadband values, while some newer metrics are based on octave band values. More modern speech metrics, such the weighted version of U_{50} called $U_{50}(A)$ [2] and STI [3], have been developed in terms of weighted octave band values. When summing octave band values, the STI and $U_{50}(A)$ require a specific octave band weighting procedure to be applied to each octave band value, in order to obtain a single broadband number that relates to speech intelligibility. When octave band values are used, the calculation of AL_{cons} is only done, by convention, in the 2 kHz octave frequency band. Therefore, the formula for predicting AL_{cons} as developed by Peutz [4], does not take into account the influence of the frequency dependency of parameters on speech intelligibility.

The two most common types of speech distortions found in room acoustics, namely reverberation and background noise, are represented in the present work by one reverberation time and one signal-to-noise ratio value. This reverberation time can be considered to be for a frequency band that represents the room acoustical conditions that are important for speech intelligibility, e.g., the 1-kHz band. A representative value for the signal-to-noise ratio can be based on the overall A-weighted levels for speech sounds and background noise.

2. Articulation for the speech intelligibility metrics

Speech metrics are physical measures of acoustical conditions for speech intelligibility which, when validated by articulation tests, allow estimates of the actual speech intelligibility experienced by listeners for a given room acoustical condition. Articulation tests are the direct form for measuring the intelligibility of speech of a communication system. In a room, these basically consist of the production of speech test material at one point in the room and listeners who try to correctly identify the speech material at some other point. The percentage of correctly identified material is the articulation score. There are different types of speech test material that result in different articulation scores. These can vary from complete sentences to single test words or even nonsense syllables, although they should all be representative of the range of sounds found in a particular language. Speech test material that includes some redundancy, such as complete sentences, can be easier to identify in adverse conditions and may lead to higher articulation scores than material such as nonsense syllables.

The speech metrics that have been proposed were usually subjectively validated in terms of different articulation tests. The main body of intelligibility scores related to STI and AL_{cons} was obtained using predominantly nonsense phonetically balanced (PB) syllables of the consonant-vowel-consonant (CVC) type for the Dutch language [4]. STI has been validated in different languages, as has been demonstrated in an evaluation with the simplified version of STI – rapid speech transmission index (RaSTI) – and calculated only in the 500 Hz and 2 kHz octave frequency bands [5]. Based on the articulation test scores, a metric-specific subjective intelligibility scale was developed to be used with STI and AL_{cons}. The Fairbanks' Rhyme Test was used to validate U_{50} [6].

Figure 1 gives the speech intelligibility obtained with the three speech metrics considered by the present study, versus reverberation time T with $L_n - L_{splm}$ as a parameter. Figure 1a is based on U_{50}, with speech intelligibility given by the Fairbanks' Rhyme Test. Figure 1b is based on STI and Figure 1c is based on AL_{cons}. Speech intelligibility according to STI and AL_{cons} is indicated by their respective subjective intelligibility scales. In this figure, the noise to-reflected-speech ratio $10^{(L_n - L_r)/10}$ was calculated according to Eq. (4), for a 300 m^3 classroom and for $L_n - L_{splm}$ values of -10, -20, and -30 dB. It can be seen that different speech metrics give different speech intelligibility results when applied to the same room acoustical conditions. For instance, for reverberation times less than about 1 s and for $L_n - L_{splm}$ equal to -20 and -30 dB, speech intelligibility according to U_{50} and given by the Fairbanks' Rhyme Test is 100%. According to STI, "excellent" speech intelligibility is only possible for reverberation times less than about 0.5 s and for $L_n - L_{splm}$ equal to -30 dB. The AL_{cons} subjective intelligibility scale indicates that "very good" speech intelligibility is possible for reverberation times less than about 1 s and for $L_n - L_{splm}$ equal to -30 dB.

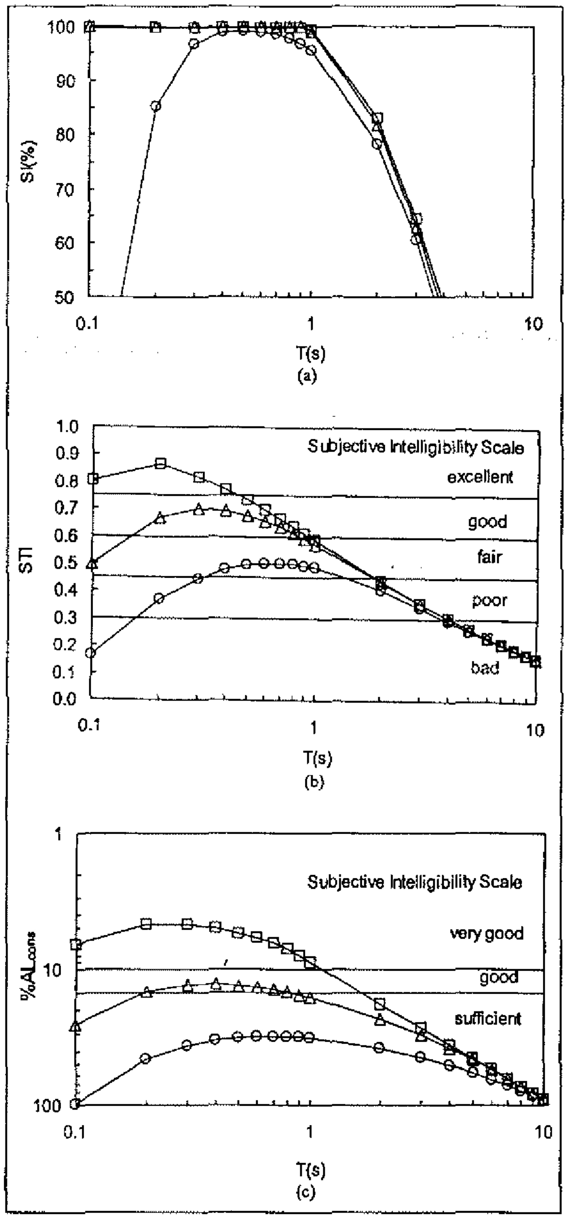

Figure 1. Speech intelligibility versus reverberation time T for a 300 m^3 classroom with L_n-L_{splm} as a parameter. The noise-to-reflected-speech ratio was calculated according to Eq. (4) with L_n - L_{splm} values of:-10 dB (o), -20 dB (Δ) and -30 dB(□). (a) speech intelligibility according to U_{50} and given by the Fairbanks' Rhyme Test; (b) speech intelligibility given by the STI subjective speech intelligibility scale; (c) speech intelligibility given by the AL_{cons} subjective speech intelligibility scale.

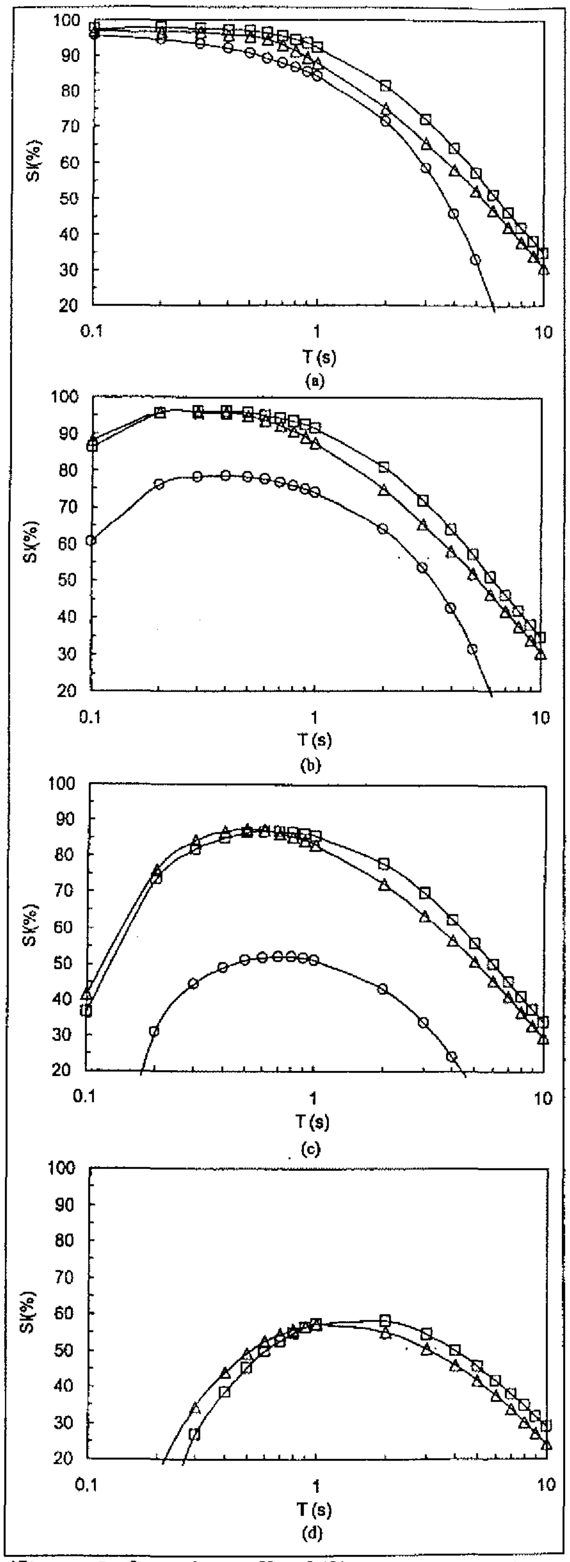

Figure 2. Speech intelligibility as a percentage SI(%), given by the Harvard PB-word test, versus reverberation time T for a 300 m^3 classroom. Speech intelligibility according to: U_{50}(Δ), STI (□) and AL_{cons} (o). (a) corresponds to the "no noise" condition. With added ambient noise, the noise-to-reflected-speech ratio was calculated according to Eq. (4) with L_n - L_{splm} values of: (b) -20 dB, (c) -10 dB and (d) 0dB., In (d), SI(%) predicted by AL_{cons} is off the bottom of the scale

Few studies have attempted to correlate different speech intelligibility metrics with the same type of articulation test. One of such studies [7] used the Harvard PB-word test to develop correlations between articulation scores and measured values of speech metrics. This articulation test is based on a set of 20 PB word lists, each of which contains 50 monosyllabic English words.

Figure 2 shows plots of speech intelligibility as a percentage %SI, given by the Harvard PB-word test, according to U_{50}, STI and AL_{cons}, versus reverberation time T for a 300 m^3 classroom. Here (a) corresponds to the "no noise" condition. With added background noise, the noise-to-reflected-speech ratio was calculated according to Equation (4), with $L_n - L_{splm}$ values of: (b) -20 dB, (c) -10 dB and (d) 0 dB.

It can be seen in Figure 2 (a) that for the "no noise" condition, speech intelligibility predictions according to STI and U_{50} are in good agreement, while AL_{cons} gives lower speech intelligibility for the same reverberation time. With added background noise, predictions according to STI and U_{50} continue to present a remarkably good agreement particularly for reverberation times less than about 1 s, whereas speech intelligibility is increasingly under-predicted by AL_{cons} as the signal-to-noise ratio decreases.

These results suggest that the effect of background noise is over-represented in the AL_{cons} predictive formula. In fact, Eq. (3) shows that AL_{cons} decreases as the signal-to-noise ratio increases up to 25 dB. As far as the effect of background noise on speech intelligibility is concerned, there is a general consensus that signal-to-noise ratios above 15 dB do not affect speech intelligibility and can be considered as "no noise" situations.

Figure 2 (a), which corresponds to the "no noise" condition, shows that speech intelligibility monotonically increases with decreasing reverberation time. However, this situation cannot be realized in real rooms because background noise is usually present. As the reverberation time decreases, the signal-to-noise ratio as given by Equation (4) decreases, which tends to deteriorate speech intelligibility. As shown in Figure 2 (b) – (d), the presence of background noise results in a relatively broad maximum, which shifts to longer reverberation times for lower signal to noise ratios.

3. Optimum acoustical conditions for speech intelligibility in classrooms

It is clear from the plots of Figures 1 and 2 that for each $L_n - L_{splm}$ value, there is a range of reverberation times for which speech intelligibility is maximized. In the case of speech intelligibility predictions according to U_{50}, Figure 1a shows that for the lower $L_n - L_{splm}$ values, there is a range of reverberation times that correspond to 100% speech intelligibility. The optimum reverberation can be defined as the reverberation time for which 100% speech intelligibility is attained with the minimum amount of sound absorption. Therefore, the optimum reverberation time is the longest reverberation time for 100% speech intelligibility.

Table I shows, for three classroom volumes and for five $L_n - L_{splm}$ values, the range of reverberation times that gives 100% speech intelligibility, and the optimum reverberation time according to U_{50}, STI and AL_{cons}. The values for STI and AL_{cons} used to obtain the reverberation times for 100% speech intelligibility, correspond to "excellent" and "very good" in their respective subjective intelligibility scales. The reverberation times listed in Table I are for a frequency band that represents the room acoustical conditions that are important for speech intelligibility, e.g., the 1-kHz band.

The range of reverberation times for 100% speech intelligibility varies according to each speech metric considered. This is due to different articulation tests used to derive each metric, and in the case of STI and AL_{cons}, also due to a rather arbitrary way of establishing the cut-off values in their respective subjective speech intelligibility scales. Both factors also influence the different values for the optimum reverberation time as given by each speech intelligibility metric.

Table I shows that for quiet classrooms, with $L_n - L_{splm}$ values in the range between -30 and -25 dB, the optimum reverberation time according to STI is 0.4s, whereas

Table I. Range of Reverberation times to achieve 100% speech intelligibility and optimum reverberation times according to U_{50}, STI and AL_{cons}, for three classroom volumes, and for five L_n - L_{splm} values. These reverberation times are for a frequency band that represents the room acoustical conditions that are important for speech intelligibility, e.g., the 1-kHz band.

$L_n - L_{splm}$ (dB)	Speech Intelligibility Metric	Range of Reverbertion Times to Achieve 100% Speech Intelligibility[a] (s)			Optimum Reverberation Time (s)		
		Classroom Volume (m^3)					
		100	300	500	100	300	500
	U_{50}	0.2-0.9	–	–	0.9	–	–
-10	STI	–	–	–	–	–	–
	Al_{cons}	–	–	–	–	–	–
	U_{50}	0.1-1.0	0.2-0.9	0.2-0.8	1.0	0.9	0.8
-15	STI	–	–	–	–	–	–
	Al_{cons}	–	–	–	–	–	–
	U_{50}	0.1-1.0	0.1-1.0	0.1-0.9	1.0	1.0	0.9
-20	STI	0.2-0.4	–	–	0.4	–	–
	Al_{cons}	0.2-0.7	–	–	0.7	–	–
	U_{50}	0.1-1.0	0.1-1.0	0.1-1.0	1.0	1.0	1.0
-25	STI	0.1-0.4	0.2-0.4	0.3	0.4	0.4	0.3
	Al_{cons}	0.1-1.1	0.1-0.8	0.3-0.4	1.1	0.8	0.4
	U_{50}	0.1-1.0	0.1-1.0	0.1-1.0	1.0	1.0	1.0
-30	STI	0.1-0.4	0.1-0.4	0.2	0.4	0.4	0.4
	Al_{cons}	0.1-1.1	0.1-1.1	0.1-1.1	1.1	1.1	1.1

[a] The values for STI and AL_{cons} used to obtain the range of reverberation times, correspond to "excellent" (˜ 0.75) and "very good" (≤10%) in their respective subjective intelligibility scales

according to U_{50} the optimum reverberation time is 1.0 s. According to AL_{cons}, the reverberation time varies from 0.4 s to 1.1 s. This analysis does not lead to precise and unambiguous indications for optimum reverberation times. However, a reverberation time of 0.4 s is the longest reverberation time that gives speech intelligibility of 100% unanimously by the three speech metrics.

With respect to the effects of background noise on speech intelligibility, Table I shows that for $L_n - L_{splm}$ values of -30 and -25 dB, speech intelligibility of 100% would be achieved according to the three speech metrics. For $L_n - L_{splm}$ equal to -20 dB, and according to STI and AL_{cons}, speech intelligibility of 100% is only achieved in 100 m^3 classrooms, whereas according to U_{50} speech intelligibility of 100% would be achieved in classrooms up to 500 m^3. For $L_n - L_{splm}$ equal to -15 dB, U_{50} indicates that speech intelligibility of 100% is still possible in classrooms up to 500 m^3, whereas according to STI and AL_{cons} speech intelligibility of 100% is no longer possible in classrooms. For $L_n - L_{splm}$ equal to -10 dB, only U_{50} indicates that speech intelligibility of 100% is still possible in 100 m^3 classrooms. The reason why U_{50} is less restrictive to the influence of background noise on speech intelligibility is due to the fact this metric was validated by the Fairbanks' Rhyme Test, which is considered a very simple type of articulation test. However, speech intelligibility of 100% is unanimously indicated by the three speech metrics with the highest $L_n - L_{splm}$ value of -25 dB. Based on this criterion, the maximum background-noise level for classrooms would be 25 dB below the voice level measured at 1 m in front of the talker. A survey of voice levels reported in the literature [1], found that an anechoic voice level of 63 dB(A) at 1 m in front of the talker can be considered as representative of the average voice level produced by teachers in classrooms for children. This would correspond to a maximum background-noise level for classrooms of 38 dB(A).

4. Summary and conclusions

The present work had the objective of determining the reverberation time and the maximum background-noise level for classrooms for best speech intelligibility conditions. This was accomplished by a comparative study of three speech intelligibility metrics, under the assumption of a diffuse sound field with ideal exponential decays. Due to the fact that the speech metrics considered by the present study were validated using different articulation tests, it is shown that they may lead to different speech intelligibility results for the same room acoustical conditions. One of the consequences was that the reverberation for best speech intelligibility conditions in classrooms could not be determined unambiguously. However, a reverberation time of 0.4 s gives speech intelligibility of 100% unanimously by the three speech metrics considered. The present results indicate that the maximum background-noise level for classrooms should be 25 dB below the voice level measured at 1 m in front of the talker.

References

[1] 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. Vol. 107, 861-875 (2000).

[2] J. S. Bradley, "Relationships Among Measures of Speech Intelligibility in Rooms," J. Audio Eng. Soc. Vol. 46, 396-405 (1998).

[3] Objective Rating of Speech Intelligibility by the Speech Transmission Index, IEC Standard 60268-16, Second Edition, 1998-03.

[4] S. R. Bistafa and J. S. Bradley, "Revisiting Algorithms for Predicting the Articulation Loss of Consonants ALcons," J. Audio Eng. Soc. Vol. 48, 531-544 (2000).

[5] T. Houtgast and H. J. M Steeneken, "A multi-language evaluation of the RaSTI-method for estimating speech intelligibility in auditoria," Acustica Vol. 54, 185-199 (1984).

[6] J. S. Bradley, "Speech Intelligibility Studies in Classrooms," J. Acoust. Soc. Am. Vol. 80, 846-854 (1986).

[7] K. D. Jacob, "Correlation of Speech Intelligibility Tests in Reverberant Rooms with Three Predictive Algorithms" J. Audio Eng. Soc. Vol.37, 1020-1030 (1989).