Predictors of speech intelligibility in rooms

Bradley, J. S.

Publisher's version / Version de l'éditeur:

Journal of the Acoustical Society of America, 80, 3, pp. 837-845, 1986-09

National Research Council Canada
Conseil national de recherches Canada

Institute for Research in Construction
Institut de recherche en construction

Predictors of Speech Intelligibility in Rooms

by J.S. Bradley

Reprinted from
Journal of the Acoustical Society of America
Vol. 80, No. 3, September 1986
p. 837-845
(IRC Paper No. 1414)

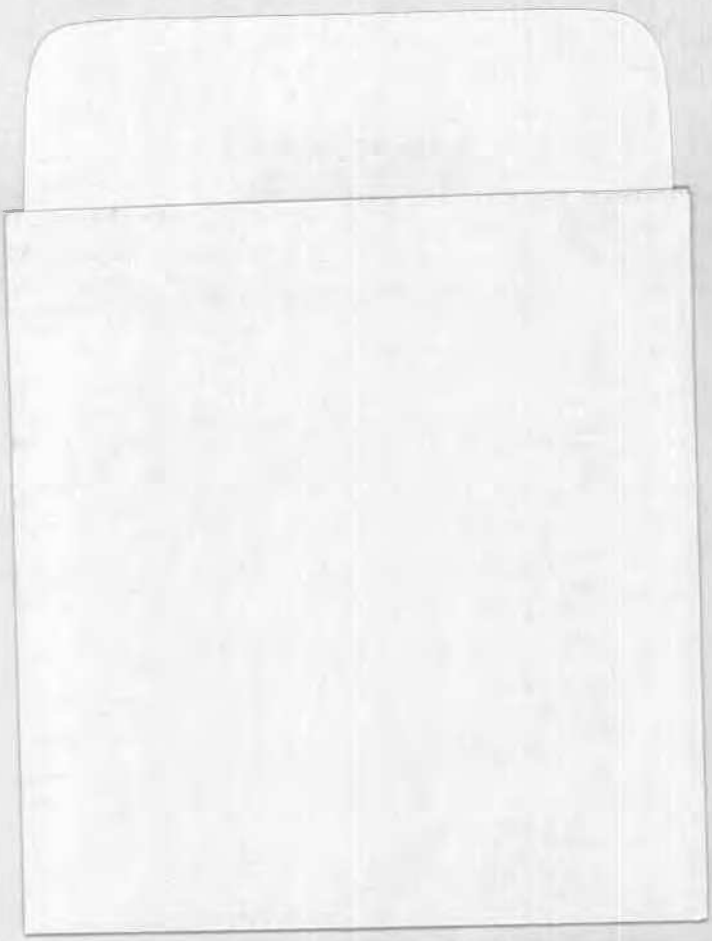
Price $2.00

NRCC 26493

Canada

RÉSUMÉ

On a comparé trois types de mesures acoustiques servant à prévoir l'intelligibilité de la parole dans des locaux de dimensions et de conditions acoustiques variées. Ces mesures portaient sur le rapport signal-bruit, sur l'indice de transmission du son tiré des fonctions de transfert de modulation, et sur les rapports son utile-son nuisible obtenus à partir des rapports son initial-son subséquent et des niveaux de parole et de bruit de fond. Pour chaque type de mesure, les formes les plus utiles ont permis d'obtenir des prévisions de précision comparable, mais c'est la mesure des rapports son utile-son nuisible basés sur un intervalle de temps initial de 0,08 seconde qui a donné la plus grande précision. Il existait des rapports étroits entre plusieurs mesures physiques, pourtant basées sur des méthodes de calcul très différentes.



Predictors of speech intelligibility in rooms

J. S. Bradley
Institute for Research in Construction, National Research Council of Canada, Ottawa, Canada K1A 0R6

(Received 23 April 1985; accepted for publication 29 April 1986)

Three different types of acoustical measures were compared as predictors of speech intelligibility in rooms of varied size and acoustical conditions. These included signal-to-noise measures, the speech transmission index derived from modulation transfer functions, and useful/detrimental sound ratios obtained from early/late sound ratios, speech, and background levels. The most successful forms of each type of measure were of similar prediction accuracy, but the useful/detrimental ratios based on a 0.08-s early time interval were most accurate. Several physical measures, although based on very different calculation procedures, were quite strongly related to each other.

PACS numbers: 43.55.Hy, 43.71.Gv

INTRODUCTION

Many rooms exist for the sole purpose of speech communication from one speaker to a group of listeners. Such rooms include a variety of sizes from small meeting rooms and classrooms to larger auditoria and theatres. The acoustical design of such rooms should be based on achieving the highest possible degree of speech intelligibility for all listeners in the room. A number of types of acoustical measures are intended to relate to the actual degree of speech intelligibility in a room, but no comprehensive comparison of these different methods has been made based on the range of conditions to be expected in rooms intended for speech. The present work compares the accuracy of various predictors of speech intelligibility from an extensive set of measurements of both physical quantities and speech intelligibility scores in a wide range of real rooms, so that the most accurate method for assessing the acoustical quality of rooms for speech can be determined.

Three types of acoustical measures were considered. The simplest type of measure is based on steady state signal-to-noise concepts. The articulation index (AI) is a well-known measure of this type.[1] In this study the overall steady-state A-weighted signal-to-noise ratio [S/N(A)] was also considered. Recently, a newer type of measure, the speech transmission index (STI) has been proposed, based on modulation transfer functions and including the effects of interfering background noise.[2,3] The third type of measure is based on the work of Lochner and Burger.[4] Although their work is now 20 or more years old, only one previous study[5] attempted to thoroughly evaluate their useful/detrimental sound ratios in a number of real rooms. In this work both the original Lochner and Burger useful/detrimental ratios and simplified forms of them that incorporated unweighted early energy sums were calculated. As the early/late-arriving sound-energy ratio for a 0.08-s early sound limit (C_{80}) has gained considerable acceptance as a correlate of subjective judgments of musical clarity,[6] it was considered likely that the same quantity, or a useful/detrimental ratio derived from it (U_{80}) would be a successful predictor of speech intelligibility scores.

I. PROCEDURE

A. The rooms

Acoustical measurements and speech intelligibility tests were performed in five rooms with volumes from 362–20 000 m^3 and 1-kHz RT values from 0.8–3.8 s. A wide range of acoustical measures was obtained from pulse recordings at 40 source–receiver combinations in the rooms, with 6–14 source–receiver positions in each room.

B. Speech tests

Speech intelligibility scores were obtained using a Fairbanks rhyme test. The procedures and the word lists were very similar to those used by Latham,[5] as initially this project was intended to replicate and expand the work of Latham. The five word lists of 50 words each were tape-recorded in an anechoic room using a male speaker and reproduced using a PSB Alpha–II loudspeaker containing a 9-cm and a 2.5-cm driver in a small box and having directional properties similar to a human speaker. For each of the 40 source–receiver combinations, tests were carried out for four different speech levels. The speech levels varied according to the background noise levels in each room. The tape recordings were calibrated by octave band and overall A-weighted level integrations over each full recording and were presented at a rate of 3.0 syllables per second. The test words were embedded in the carrier phrase, "Word number ___ is ___ . Write that down please." For each receiver position, a group of nine subjects were seated as close as possible to the actual receiver location, and the scores of the nine subjects were averaged to obtain one score representative of that receiver location. With nine subjects at each of the 40 source–receiver combinations repeated at four speech levels, a total of 1440 individual speech-intelligibility tests were performed.

C. Subjects

Subjects received only a few minutes of instruction prior to the test and no hearing tests were performed. However, subjects who reported a known hearing impairment were not used. Subjects were thus not carefully selected listeners, but

were more representative of general listening audiences without obvious hearing-handicapped subjects. Different groups of subjects were used to test each room, with as many as 45 subjects used in one room. Subjects varied in age from approximately 16-year-old high school students to working age adults. No younger children or retirement age adults were included.

D. Acoustical measurements

For each source–receiver combination, pulses were recorded using pistol shots. From these recorded pulses, the early/late-arriving sound ratios, early decay times (EDT), and the conventional RT were calculated in octave bands from 125–8000 Hz. The calculation of the measures from pulses and the interrelation of the basic physical measures were considered in a previous paper.[7]

Background noise level recordings, as well as the pulse measurements, were made at each receiver location. The 1-min integrations of the octave band and overall A-weighted background noise were made from these recordings. By calibration of the reproduction system and the recorded test tapes, the speech levels were determined in terms of the integrated level of the complete test tape, at a distance of 1.0 m, in an anechoic environment. From the measured source–receiver distances, RT values, and loudspeaker directivity factors, the long-time average speech levels were calculated in octave bands at each receiver position using

$$\mathrm{SPL} = \mathrm{SWL} + 10\log(Q/4\pi r^2 + 4/A) + 0.1, \text{ dB}, \quad (1)$$

where SWL is the sound power level of the source, Q is the directivity factor of the loudspeaker source and is a function of frequency and angle, r is the source–receiver distance in meters, and A is the total absorption in the room in square meters. The absorption was calculated from the measured RT values, using the Sabine reverberation equation. Measuring the long-time average speech levels would have permitted increased accuracy in the speech levels at each receiver, but the procedure used better indicates the uncertainty that would be expected in future attempts to predict speech intelligibility scores from the results of the present studies. That is, in such future situations, speech levels would probably not be measured at each receiver. One would want to predict speech intelligibility scores from known or estimated source levels and the properties of the room.

E. Calculation of predictors

Three types of acoustical measure were considered as predictors of speech intelligibility scores: Steady-state signal-to-noise measures; measures derived from early/late-arriving sound ratios, speech, and background noise levels; and STI values derived from modulation transfer functions, speech, and background noise levels. The articulation index (AI) is the best known of the first type and standardized procedures for its calculation are available.[1] First, 12 dB is added to the long-time average signal-to-noise ratios in each standard octave band from 125–8 kHz. Each signal-to-noise ratio is multiplied by a weighting factor and the weighted ratios are then summed to produce an AI value between 0 and 1.0. The + 12 dB is intended to represent the difference between the long-time average speech levels and the peak levels. In this study a simple overall A-weighted signal-to-noise ratio [S/N(A)] was calculated for each speech level at each receiver.

The STI values were calculated from pulses as described in an earlier paper,[7] ignoring the effects of background noise levels. The octave band weightings suggested by Steeneken and Houtgast[3] were used, and the effect of the steady state signal-to-noise ratios was combined with the modulation transfer functions as follows:

$$m_n(\omega) = m(\omega) \cdot [1 + 10^{(-\mathrm{SN}/10)}]^{-1}, \quad (2)$$

where $m(\omega)$ is the modulation transfer function with no interfering noise and $m_n(\omega)$ is the corresponding value with noise. The SN is the long-time average steady-state signal-to-noise ratio in decibels. The STI values were calculated both with and without noise, following the steps outlined by Houtgast *et al.*[2]

Lochner and Burger[4] introduced the concept of the ratio of useful/detrimental sound energy that was intended to relate to speech intelligibility scores. The useful energy was a weighted sum of the energy arriving in the first 0.095 s after the arrival of the direct sound. The detrimental energy was the later-arriving energy from the speech source, plus the background noise energy in the room. Such a measure is essentially an early/late sound ratio, with the background sound energy added to the late-arriving sound. Various other forms of this measure could be calculated from early/late sound ratios. Such measures would be less complicated to calculate than those using the weighting procedure proposed by Lochner and Burger, which is quite complicated and requires the identification of individual reflections in the pulse response. The Lochner and Burger form of early/late ratio is referred to as C_{95} and is given by the following equation:

$$C_{95} = 10\log\left(\int_0^{0.095} \alpha \cdot p^2(t)\,dt \Big/ \int_{0.095}^{\infty} p^2(t)\,dt\right), \text{ dB}, \quad (3)$$

where α is the fraction of the energy of an individual reflection that is integrated into the useful early energy sum. In the present work the early sound energy in each block of 12 points of the pulse response (sampled at 22627 Hz), was summed to create intermediate energy sums. Each of these sums was then treated as an individual reflection. By Lochner and Burger's technique, the weighting of each individual refelection is determined by its relative amplitude (RA) (relative to the direct sound) and by its arrival time (T) after the direct sound (as seen in Fig. 10 in Ref. 4). The curves of this figure were approximated in the present work using the following relationships for α, the fraction of the energy integrated into the useful early energy:

$$\alpha = A + B \cdot T, \quad (4)$$

where $A = 2.30 - 0.600 \cdot \mathrm{RA}^{0.7}$ and $B = -0.0248 + 0.00177 \cdot \mathrm{RA}^{1.35}$. The variable T is the time of arrival after the direct sound in milliseconds, and α values are limited so that they fall between 0 and 1.0. The relative total early useful energy was then obtained by adding up the block energy sums weighted by the appropriate α values.

The calculated relative early useful energy and late ener-

gy sums were used to find the fraction of the total speech energies that were useful or detrimental. If SL and BL are the steady-state long-term rms speech and background levels, E_{SL} and E_{BL} are the related total speech and background energies:

$$E_{SL} \propto 10^{(SL/10.0)}$$

$$E_{BL} \propto 10^{(BL/10.0)}.$$

The useful early energy is then

$$\text{Useful} = [E_e/(E_e + E_l)] \cdot E_{SL}$$

$$= [C_{te}/(C_{te} + 1)] \cdot E_{SL}, \quad (5)$$

where E_e and E_l are the relative early and late energy sums from the uncalibrated pulse recordings. The variable C_{te} is the linear early/late ratio with an early time limit (te). Similarly, the detrimental energy is given by

$$\text{Detrimental} = [1/(C_{te} + 1)] \cdot E_{SL} + E_{BL}. \quad (6)$$

By dividing Eq. (5) by Eq. (6), and after some simplification, an expression for a useful/detrimental ratio is obtained for an early sound limit (te),

$$U_{te} = C_{te}/[1 + (C_{te} + 1) \cdot E_{BL}/E_{SL}]. \quad (7)$$

Thus useful/detrimental sound ratios can be calculated from the corresponding early/late ratio (C_{te}) and the ratio of background noise to speech energies.

Early/late ratios were calculated for 0.035-, 0.050-, 0.080-, and 0.095-s early sound limits and are referred to as C_{35}, C_{50}, C_{80}, and C_{95}. Of course, C_{95} is different from the others in that the early energy sum involves the weighting of the individual reflections described above. From these early/late ratios, useful/detrimental sound ratios were then calculated for the corresponding early time limits and are referred to as U_{35}, U_{50}, U_{80}, and U_{95}. All measures were calculated in the seven octave bands from 125–8 kHz.

II. RESULTS

A. Signal-to-noise measures

Figure 1 plots the results of 160 mean speech intelligibility test scores versus AI values. The line shown on the figure is the result of fitting a third-order polynomial to the data.

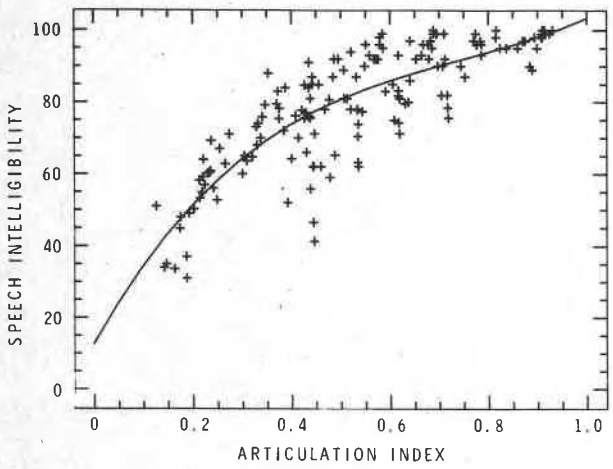


FIG. 1. Measured speech intelligibility scores versus AI values and best-fit third-order polynomial.

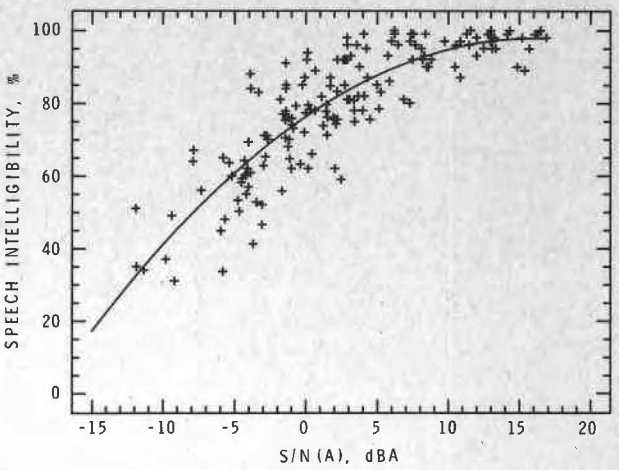


FIG. 2. Measured speech intelligibility scores versus S/N(A) values and best-fit third-order polynomial.

As many of the relationships between predictors are clearly not linear, in all cases, third-order polynomials were fitted to the data and the resulting multiple correlation coefficients were calculated along with the standard error (SE) of the estimate. This uniformity of approach permits direct comparisons between all correlations. In the case of AI values, the multiple correlation coefficient was 0.847 and the SE was $\pm 9.1\%$.

Similar results are shown for the S/N(A) in Fig. 2, again along with the best-fit third-order polynomial. Here the multiple correlation coefficient was 0.862 and the SE was $\pm 8.7\%$. Thus the simple S/N(A) was a slightly better predictor of the present intelligibility scores.

The ANSI standard for the calculation of AI values[1] inculdes a correction for RT value. Compound predictors were created based on combinations of AI and RT values. As EDT relates more closely to subjectively perceived decay time, attempts were made to create compound measures that included it too. Table I summarizes the results for combinations of third-order polynomials of both AI and S/N(A), combined with RT and EDT values. All four results produced similar multiple correlation coefficients, and SE values. The SE values ranged from ± 7.2–7.5%, and were slightly smaller for combinations with S/N(A).

B. Modulation transfer function measures

STI values were calculated both with and without the influence of background noise levels. Without background noise levels the multiple correlation coefficient was 0.525 and the SE was $\pm 14.6\%$. Figure 3 shows measured speech intelligibility scores versus STI values that included the effects of background noise levels. The best fit third-order polynomial is also shown; the multiple correlation coefficient was 0.866 and the SE $\pm 8.6\%$. Thus, for the present data, the STI is of similar prediction accuracy to the A-weighted signal-to-noise ratio. Steeneken and Houtgast[3] reported a standard deviation about their best-fit relationship of $\pm 5.6\%$ for PB-word scores versus STI values. Their data comprised 167 points, some of which included reverberation.

TABLE I. Multiple regression results, compound predictors of speech intelligibility.

Independent variables		Regression coefficients						
X1	X2	X1	$X1^2$	$X1^3$	X2	Constant	R	SE
AI	RT(1 kHz)	273.5	−316.0	134.3	−5.744	18.40	0.901	7.5
AI	EDT(1 kHz)	273.3	−319.5	137.8	−5.326	17.60	0.900	7.5
S/N(A)	RT(1 kHz)	2.638	−0.1222	0.001282	−5.702	88.82	0.908	7.2
S/N(A)	EDT(1 kHz)	3.603	−0.1205	0.001330	−5.267	89.51	0.906	7.3

C. Early/late ratio measures, complete data

Early/late ratios, decay times, and useful/detrimental ratios were all considered as predictors of speech intelligibility scores. In all cases, third-order polynomials were fitted to the data relating speech intelligibility scores and the octave band predictors. Table II(a) gives the resulting multiple correlation coefficients and Table II(b) gives the corresponding SE values. All useful/detrimental ratios correlated much more strongly with speech intelligibility scores than either the early/late ratios or the decay times. In tests such as those in the present study, where the steady-state signal-to-noise ratio was deliberately manipulated as part of the experiment, early/late ratios and decay times are inadequate predictors, as they only indicate one aspect of the problem. One must assume that much of the scatter is due to large differences in signal-to-noise ratio between data points. The useful/detrimental ratios were much better predictors of speech intelligibility scores, with the highest correlations in the 2-, 4-, and 8-kHz octave bands. The particular early time limit did not seem to have a strong influence on the multiple correlation coefficients, but the U_{80} values, with a 0.08-s early time limit, produced the strongest correlations. The U_{95} values, using the Lochner and Burger weighting factors in the calculation of the early useful sound energy, were not superior to the other less complex measures. The multiple correlations in the lowest four octave bands tended to be a little smaller than in the highest three bands.

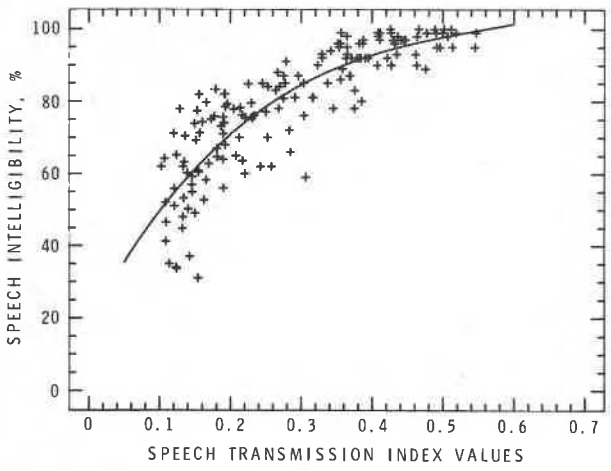


FIG. 3. Measured speech intelligibility scores versus STI values (including background noise levels) and best-fit third-order polynomial.

The SE values in Table II(b) show that speech intelligibility scores could be predicted with a standard error as small as ± 6.8%. This minimum SE value occurred for the 125-Hz and 2-kHz U_{80} values. This is smaller than the SE value for AI values, S/N(A) values, compound AI and S/N(A) predictors with decay times, or for STI values as predictors of speech intelligibility scores. Figure 4 plots speech intelligibility scores versus U_{80} values and also shows the best-fit third-order polynomial. This best-fit curve, for predicting speech intelligibility (SI), is given by the following equation:

$$SI = 1.219 \cdot U_{80} - 0.02466 \cdot U_{80}^2 + 0.00295 \cdot U_{80}^3 + 95.65. \quad (8)$$

Figure 5 is a similar plot in terms of U_{95} values, and includes both the best-fit third-order polynomial to the measured data and Latham's best-fit curve to his data. The best-fit line on this figure is given by

$$SI = 0.7348 \cdot U_{95} - 0.09943 \cdot U_{95}^2 - 0.0005457 \cdot U_{95}^3 + 197.39. \quad (9)$$

Latham's curve differs a little from the best fit to the present data. This difference can be attributed to a number of differences in procedures between Latham's work and the present study. (1) Latham quotes speech levels in terms of long-time mean sound levels, whereas in the present study, speech levels are the more usual long-time energy average levels. (2) Latham measured the attenuation of speech levels from the source to each receiver position, but in the present study these effects were only calculated. (3) Latham used a mean maximum PNC value as a measure of background levels obtained from sound level meter readings. In the present study, 1-min energy average A-weighted background levels were used. (4) While Latham used one typical value to represent the background levels in a hall, in the present, background levels were measured at each receiver position. Latham's use of mean maximum PNC background levels and mean speech levels would produce lower values than the energy average levels in the present study. However, the difference between the two quantities, which influences the final useful/detrimental ratios, would probably be similar to the values in the present research. (5) Finally, Latham used the Schroeder integrated impulse response technique to obtain a smooth decay curve (equivalent to the ensemble average of a large number of decays from steady-state noise) as an impulse response from which to identify individual reflections, and then applied Lochner- and Burger-type weighting

TABLE II.

Independent variable	Octave-band frequency, Hz 125	250	500	1000	2000	4000	8000
(a) Multiple correlation coefficients from third-order prediction equations of speech intelligibility scores							
C_{35}	(0.120)ns[a]	(0.218)ns	0.369	0.421	0.516	0.462	0.474
C_{50}	(0.189)ns	0.337	0.404	0.475	0.567	0.541	0.500
C_{80}	(0.172)ns	0.329	0.442	0.485	0.582	0.574	0.524
C_{95}	0.236	0.234	0.391	0.473	0.492	0.532	0.480
RT	0.236	0.315	0.580	0.609	0.642	0.571	0.606
EDT	0.268	0.371	0.593	0.587	0.624	0.600	0.534
U_{35}	0.866	0.923	0.829	0.870	0.901	0.903	0.903
U_{50}	0.884	0.913	0.849	0.874	0.909	0.911	0.913
U_{80}	0.918	0.907	0.878	0.898	0.919	0.915	0.913
U_{95}	0.849	0.870	0.820	0.874	0.898	0.913	0.911
(b) Standard errors, percent							
C_{35}	(17.1)ns	(16.8)ns	16.0	15.6	14.7	15.2	15.1
C_{50}	(16.8)ns	16.2	15.7	15.1	14.2	14.5	14.9
C_{80}	(16.9)ns	16.2	15.4	15.0	14.0	14.1	14.6
C_{95}	16.7	16.7	15.8	15.1	14.9	14.5	15.1
RT	16.7	16.3	14.0	13.6	13.2	14.1	15.1
EDT	16.5	16.0	13.8	13.9	13.4	13.7	14.5
U_{35}	8.6	6.6	9.6	9.5	7.4	7.4	7.4
U_{50}	8.0	7.0	9.1	8.4	7.2	7.1	7.0
U_{80}	6.8	7.2	8.2	7.5	6.8	6.9	7.0
U_{95}	9.0	8.5	9.8	8.4	7.6	7.0	7.1

[a] ns = not significant, $p \leq 0.05$.

functions to sum the early useful energy. He did not describe how he calculated the weightings of each early reflection. His decays would not be the same as the original impulse response in that they would not contain spikes representing individual reflections. Latham's data, although obtained from seven different rooms, included 95 sets of test data with a more limited range of reverberation times. It is not possible to explain in detail how each of these procedural differences contributed to the final results of each study, but there were enough significant differences to lead to the overall differences, seen from the curves in Fig. 5.

D. Early/late measures, high signal-to-noise ratios

The poor correlations between speech intelligibility scores and early/late ratios or decay times were assumed to be due to the large variance attributed to variations in signal-to-noise ratios. If the signal-to-noise ratio is high enough, then the background noise would have only a minor effect on speech intelligibility. Accordingly a subset of the data was created where S/N(A) ratios were all greater than 10.0 dB. This subset contained 43 cases. For the 500-Hz and 1-kHz octave bands, the data of this subset were then considered as before, by fitting third-order polynomials of the predictor variables to obtain best-fit curves with speech intelligibility scores. Table III gives the resulting multiple correlation coefficients. When the detrimental effects of background noise are reduced in this way, the multiple correlation coefficients between speech intelligibility scores and early/late ratios

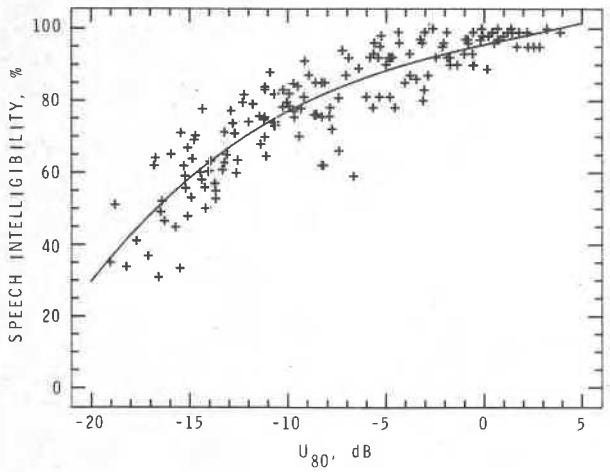


FIG. 4. Measured speech intelligibility scores versus 1-kHz U_{80} values and best-fit third-order polynomial.

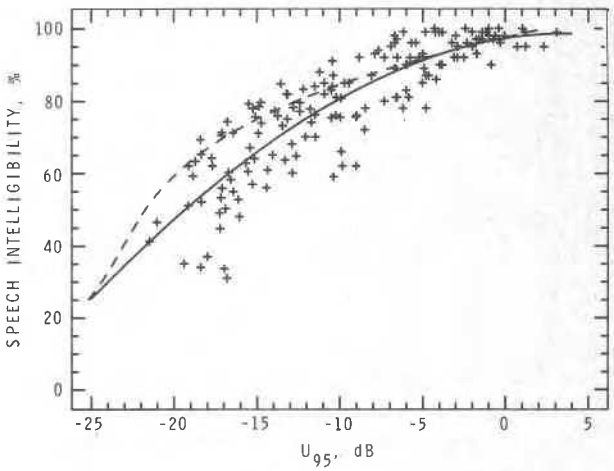


FIG. 5. Measured speech intelligibility scores versus 1-kHz U_{95} values, best-fit third-order polynomial (solid lines) and Latham's best-fit curve (dashed line).

TABLE III. Multiple correlation coefficients from third-order prediction equations, S/N(A)≥10 dB.

Independent variable	Octave-band frequency, Hz	
	500	1000
C_{35}	0.706	0.741
C_{50}	0.712	0.725
C_{80}	0.499	0.593
C_{95}	0.547	0.665
RT	0.401	0.537
EDT	(0.185)ns[a]	0.478
U_{35}	0.700	0.813
U_{50}	0.776	0.716
U_{80}	0.567	0.683
U_{95}	0.583	0.767

[a] ns = not significant, $p < 0.05$.

were increased and were similar to those with useful/detrimental ratios. Thus speech intelligibility scores are influenced by the strength of the early-arriving sound relative to the strength of the later-arriving sound. However, the adverse effects of inadequate signal-to-noise ratios seem to be more critical, and have larger effects on the resulting speech intelligibility. This result may be partly due to the nature of the present experiment, where speech source level was deliberately varied over a wide range, but at relatively low levels, compared to typical speech. However, even for the limited case described above, where all S/N(A) values were greater than 10.0 dB, speech intelligibility scores were still significantly related to the S/N(A) values.

E. Early/late measures, multiple band combinations

The AI values, S/N(A) ratios, and STI values are broadband measures, incorporating information from the seven octave bands from 125–8000 Hz. Attempts were made to calculate useful/detrimental ratio measures including information from three octave bands. Two different sets of three octave bands were tried and two types of summation were considered. The octave bands included 500, 1000, and 2000 Hz in one case and 1000, 2000, and 4000 Hz in the other case. For one type of summation, the useful/detrimental ratio values from the three octave bands were simply arithmetically summed, while in the other type of summation, the useful/detrimental ratios were converted to pressure-squared values, added as energies, and converted back to decibels. The resulting multiple correlation coefficients and associated SE values are given in Table IV. The simple arithmetic summations produced more accurate predictions than the logarithmic summations and the sums that include the 4-kHz octave band produced more accurate predictions than those that included the 500-Hz octave band. The results were disappointing in that the created multiple band predictors were no more accurate than the best single band predictors.

Broadband useful/detrimental ratios were also created using an A-weighting technique and also using the octave band weighting factors of the AI procedure. Octave band U_{80} values were A-weighted and added together as energies to obtain an A-weighted U_{80} value. Multiple regression analysis was then used to find the best third-order polynomial of the A-weighted U_{80}, as a predictor of speech-intelligibility scores. A multiple correlation coefficient of 0.905 was obtained with an SE of $\pm$ 7.3%. Again, this was no better than several single-octave band U_{80} values. When a broadband measure was formed by weighting the linear pressure-squared ratios from U_{80} values with the octave band weightings used in the AI procedure, the result was a less successful predictor. The result of the multiple regression analysis, using this new measure as a third-order polynomial predictor, was a multiple correlation coefficient of 0.827 with an SE of $\pm$ 9.7%.

F. Inter-relation of physical measures

In a previous study, various early/late ratios and decay times were quite strongly correlated.[7] A consideration of the relationships between measures, not considered in the previous study, can assist in a further understanding of the present results. Accordingly, Fig. 6 plots C_{50} values versus mea-

TABLE IV. Multiple regression analyses, mutliple band predictors.

Summation type	Summed independent variables			R	SE
Arithmetic	U_{35}(500 Hz)	U_{35}(1 kHz)	U_{35}(2 kHz)	0.883	8.1
	U_{50}(500 Hz)	U_{50}(1 kHz)	U_{50}(2 kHz)	0.889	7.9
	U_{80}(500 Hz)	U_{80}(1 kHz)	U_{80}(2 kHz)	0.906	7.3
	U_{95}(500 Hz)	U_{95}(1 kHz)	U_{95}(2 kHz)	0.884	8.0
Arithmetic	U_{35}(1 kHz)	U_{35}(2 kHz)	U_{35}(4 kHz)	0.903	7.4
	U_{50}(1 kHz)	U_{50}(2 kHz)	U_{50}(4 kHz)	0.909	7.2
	U_{80}(1 kHz)	U_{80}(2 kHz)	U_{80}(4 kHz)	0.916	6.9
	U_{95}(1 kHz)	U_{95}(2 kHz)	U_{95}(4 kHz)	0.910	7.1
Logarithmic	U_{35}(500 kHz)	U_{35}(1 kHz)	U_{35}(2 kHz)	0.851	9.0
	U_{50}(500 kHz)	U_{50}(1 kHz)	U_{50}(2 kHz)	0.854	8.9
	U_{80}(500 kHz)	U_{80}(1 kHz)	U_{80}(2 kHz)	0.879	8.2
	U_{95}(500 kHz)	U_{95}(1 kHz)	U_{95}(2 kHz)	0.853	9.0
Logarithmic	U_{35}(1 kHz)	U_{35}(2 kHz)	U_{35}(4 kHz)	0.894	7.7
	U_{50}(1 kHz)	U_{50}(2 kHz)	U_{50}(4 kHz)	0.893	7.7
	U_{80}(1 kHz)	U_{80}(2 kHz)	U_{80}(4 kHz)	0.901	7.5
	U_{95}(1 kHz)	U_{95}(2 kHz)	U_{95}(4 kHz)	0.904	7.4

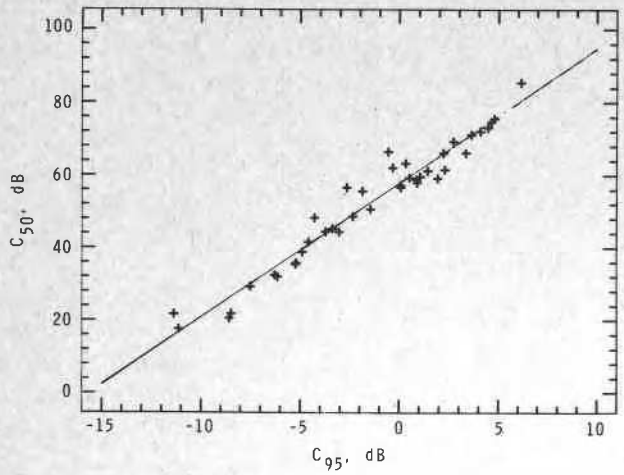


FIG. 6. Measured C_{50} values versus C_{95} values, both at 1 kHz.

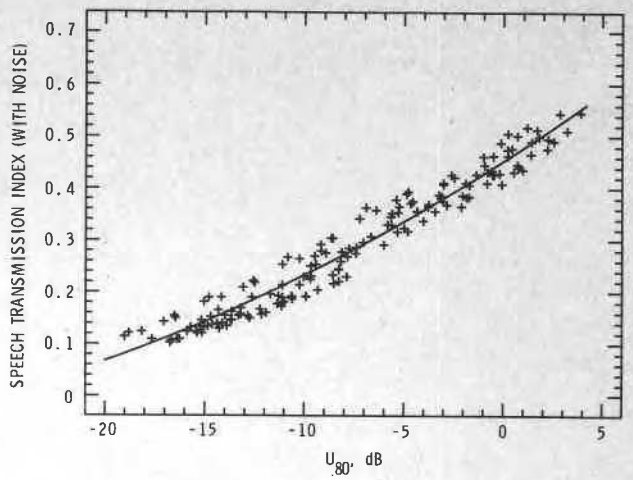


FIG. 8. Measured STI values including the effects of background noise versus U_{80} values at 1 kHz and the best-fit third-order polynomial.

sured C_{95} values. In spite of the quite different derivations of these two measures, the resulting values are remarkably similar. Similarly, U_{50} and U_{95} values were nearly equal in value. This explains why correlations with U_{50} and U_{95} values produced such similar results, and suggests that the complicated weighting procedure involved in the calculating of C_{95} values is not necessary, particularly since the other useful/detrimental ratios were at least as accurate as U_{95}, as predictors of speech-intelligibility scores.

Although STI has been strongly promoted as a predictor of speech intelligibility, the present results suggest that it is similar, but slightly inferior, to useful/detrimental ratios. Figures 7 and 8 compare values of C_{80} and no-noise STI and values of U_{80} and STI with noise, respectively. Figure 7 shows that no-noise STI values are closely related to C_{80} values: They are essentially another way of measuring almost the same property of a room. Figure 8 demonstrates that even with the added complexity of added background noise, STI values with noise and U_{80} values are very closely related. Of course the relationship is not exact, but the fact that U_{80} values are more highly correlated with speech intelligibility scores suggests that some of these small differences make U_{80} a more accurate predictor.

III. OPTIMUM CONDITIONS FOR SPEECH

The results of the present study can be used to deduce acoustical design goals for ideal conditions for speech, in terms of specific acoustical measures. Of the various measures considered in this paper, a combination of the A-weighted signal-to-noise ratio, with the RT, is the simplest, and involves only measurements that are readily made with commonly available acoustical measuring equipment. Figure 9 plots curves corresponding to the regression coefficients for this case from Table I for RT values of 1, 2, 3, and 4 s. Due to the nature of the rhyme test that was used, a satisfactory minimum speech-intelligibility score must be very high.

Latham[5] suggests that speech intelligibility scores should be 97% for excellent speech intelligibility. In this study, the point where the best-fit curves are closest to 100% speech intelligibility is used as an ideal design goal. Using this criterion, the results of Fig. 9 are in agreement with conventional acoustical wisdom: For optimum conditions, an RT of slightly less than 1.0 s and a signal-to-noise ratio of at least 15 dBA are required.

From the optimum signal-to-noise ratios, known speech

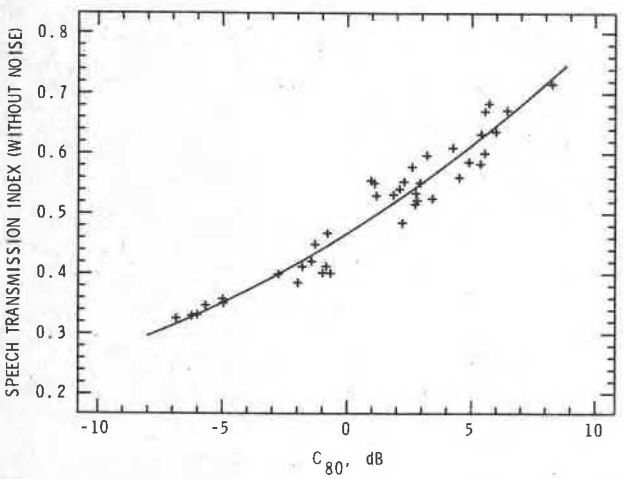


FIG. 7. Measured C_{80} values at 1 kHz versus STI values excluding the effects of background noise.

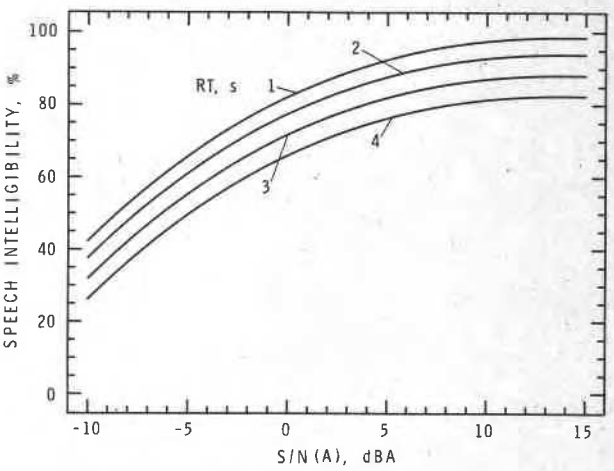


FIG. 9. Best-fit curves of speech intelligibility versus overall signal-to-noise ratio for RT values of 1, 2, 3, and 4 s.

TABLE V. Calculation of worst-case room effects.

Volume, m^3	RT, seconds	Source–receiver distance, m	Room effect, dB
300	0.5	8	− 2
1000	0.7	8	− 6
3000	0.9	16	− 10
10 000	1.1	16	− 14
30 000	1.3	16	− 17

levels, and the reduction in levels between source and receiver positions in a room, one can also estimate maximum acceptable background noise levels. Pearsons *et al.*[8] found the mean and standard deviation of speech levels for groups of people at five different levels of vocal effort. These overall A-weighted source levels were slightly lower for female speakers and so the data for females should be used as a worst case. The levels of vocal effort were: casual, normal, raised, loud, and shout. Presumably, the level of vocal effort would vary with room size. In smaller rooms, "normal" vocal effort would be expected, whereas in larger rooms, a "raised" vocal effort would be more likely. The room effect, the reduction in long-term rms speech levels from a one meter source position to the various receiver positions, would also vary from room to room. This room effect increases with decreasing RT, and with increasing source–receiver distance. By calculating these room effect values, using Eq. (1) for a large number of combinations of room size, RT, and source–receiver distance, frequently occurring worst-case room effects were estimated for various room sizes. Table V summarizes the calculation of the worst-case room effects, which varied from − 2 to − 17 dB for room volumes of 300–30 000 m^3. To calculate optimum background levels for very good speech conditions, the female speech source levels from Ref. 8 were reduced by one standard deviation to include the majority of speakers. They were then further reduced by the ideal minimum S/N(A) of 15 dBA, and also by the worst case room effect for each size of room. These calculations, summarized in Table VI, suggest ideal maximum background levels of between 27–34 dBA. For smaller rooms of 300–1000 m^3, ideal maximum background levels of 30–34 dBA were obtained. For medium sized rooms of 3000 m^3, background levels of 26 and 34 dBA were obtained, depending on the level of vocal effort. In large rooms with volumes of 10 000–30 000 m^3, ideal maximum background levels of 27–30 dBA were obtained, depending on the room size. Beranek[9] suggests maximum background levels of 30 dBA for large theatres and auditoria, and 42 dBA for small theatres and auditoria. Thus the present ideal-maximum values for 3000–30 000 m^3 rooms are quite close to Beranek's large auditoria values, but the present ideal-maximum values for rooms of 300–1000 m^3 are more conservative than Beranek's recommended small-auditoria values.

One would similarly like to estimate minimum U_{80} values to provide a high-quality speech environment. From the best-fit curve of Fig. 4, a mean score of 100% is reached for a U_{80} value of + 4 dB for these 1-kHz results. When background noise levels are very low, the C_{80} will be equal to the U_{80}. In this special case, one would thus require a minimum C_{80} of + 4 dB at 1 kHz, for very good speech conditions. From the results of a previous paper,[7] this optimum C_{80} value would correspond to a C_{50} of approximately + 1 dB or to a Deutlichkeit value of 0.56. (Deutlichkeit is the linear ratio of the early-arriving sound energy in the first 50 ms to the total energy.)

For the more realistic situations with some non-negligible background noise, it is of interest to understand how C_{80} values and background noise levels combine to produce U_{80} values. Accordingly, Fig. 10 plots calculated 1-kHz U_{80} values versus 1-kHz C_{80} values for various A-weighted overall signal-to-noise ratios. A 1-kHz U_{80} of + 4 dB can be obtained by various combinations of C_{80} and overall signal-to-noise ratio. For example, a 1-kHz U_{80} of + 4 dB could be achieved by a combination of a 1-kHz C_{80} of + 4 dB and a signal-to-noise ratio of 25 dBA, or with a 1 kHz C_{80} of 6.3 dB and a background signal-to-noise ratio of 15 dBA.

To fully utilize U_{80} values in the design of rooms for speech, it is necessary to predict C_{80} values in rooms. Recent work by Barron[10] has considered the problem of estimating C_{80} values in rooms. Further work is now required along these lines, combined with studies to gain more familiarity with the use of U_{80} values, and to determine what values represent satisfactory conditions in particular rooms.

IV. CONCLUSIONS

The results of this work suggest that several methods of almost equivalent prediction accuracy can be used for estimating expected speech intelligibility scores obtained using a Fairbanks rhyme test. The simplest approach, both in terms of performing calculations or in terms of making measurements, would be to obtain the A-weighted signal-to-noise ratio [S/N(A)] and the 1 kHz RT value and use the regression coefficients in Table I to form a prediction equation, as illustrated in Fig. 9, to estimate the expected speech-intelligi-

TABLE VI. Calculation of optimum background levels.

Room volume, m^3	Vocal effort	Speech source level, dBA	Standard deviation, dBA	Optimum S/N(A), dBA	Room effect, dB	Optimum background level, dBA
300	Normal	55	− 4	− 15	− 2	34
1000	Normal	55	− 4	− 15	− 6	30
3000	Normal	55	− 4	− 15	− 10	26
3000	Raised	63	− 4	− 15	− 10	34
10 000	Raised	63	− 4	− 15	− 14	30
30 000	Raised	63	− 4	− 15	− 17	27

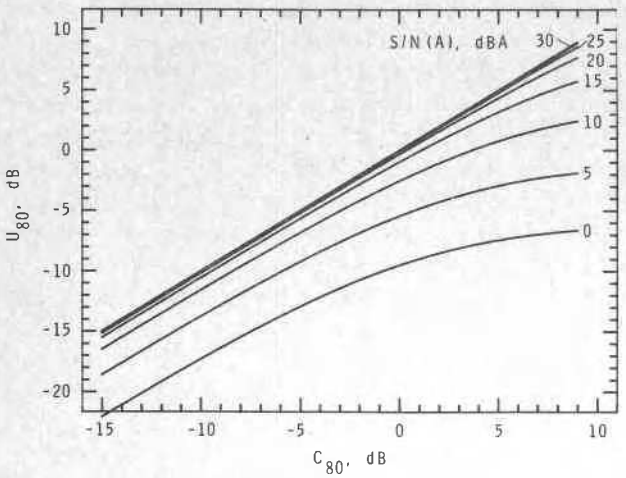


FIG. 10. Calculated U_{80} vs C_{80} at 1 kHz for overall signal-to-noise ratios of 0, 5, 10, 15, 20, 25, and 30 dBA.

bility score. With the present data this method was almost as accurate as useful/detrimental sound ratios, and more accurate than STI values.

The present data may overemphasize the importance of the signal-to-noise ratio, and may influence the success of the combination of S/N(A) and RT (1 kHz) as a good predictor of speech intelligibility. It is safer and more generally reliable to use the 0.08-s useful/detrimental ratio (U_{80}) as the preferred predictor of speech intelligibility. This has the added advantage that C_{80} values, from which U_{80} values are derived, are useful in rooms where music is also to be performed. The use of U_{80} values based on early/late ratios is also desirable, as these values are easily related to the fundamental physical quantities involved. One can readily grasp that increased or stronger early reflections lead to larger U_{80} values and hence, to increased speech intelligibility scores, and such increases can be observed directly from the pulse response in a room as displayed on an oscilloscope. The changes to a room can be directly related to changes in the degree of speech intelligibility in the room. The STI seems to be a somewhat equivalent, although a little less accurate, predictor. Unfortunately, it requires a separate set of measuring equipment to make measurements of STI values from steady-state test signals, or a quite powerful computer and associated software if values are to be calculated from pulses, as in this study.

ACKNOWLEDGMENTS

The authors would like to thank Reina Lamothe for her considerable efforts in assisting with programming and processing of the data, and the National Arts Center for their cooperation and assistance. This paper is a contribution from the Institute for Research in Construction, National Research Council of Canada.

[1]"Methods for the Calculation of the Articulation Index," ANSI Standard S3.5-1969 (American National Standards Institute, New York, 1969).
[2]T. Houtgast, H. J. M. Steeneken, and R. Plomp, "Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. I. Genera Room Acustics," Acustica **46**, 60–72 (1980).
[3]H. J. M. Steeneken and T. Houtgast, "A Physical Method for Measuring Speech Transmission Quality," J. Acoust. Soc. Am. **67**, 318–326 (1980).
[4]J. P. A. Lochner and J. F. Burger, "The Influence of Reflections on Auditorium Acoustics," J. Sound Vib. **1**, 426–454 (1964).
[5]H. G. Latham, "The Signal-to-Noise Ratio for Speech Intelligibility—An Auditorium Acoustics Design Index," Appl. Acoust. **12**, 253–320 (1979).
[6]W. Reichardt and U. Lehman, "Optimierung von Raumeindruck und Durchsichtigkeit von Musikdarbietungen durch Auswertung von Impulsshalltests," Acustica **48**, 174–185 (1981).
[7]J. S. Bradley, "Auditorium Acoustics Measures from Pistol Shots," J. Acoust. Soc. Am. **80**, 199–205 (1986).
[8]K. S. Pearsons, R. L. Bennett, and S. Fidell, "Speech Levels in Various Noise Environments," Bolt Beranek and Newman Inc., Report to U. S. EPA (Canoga Park, CA, May 1977), pp. 270–053.
[9]L. L. Beranek, *Noise and Vibration Control* (McGraw-Hill, New York, 1971), p. 585.
[10]M. Barron and L.-J. Lee, "Energy Relations in Concert Auditoria, I." J. Acoust. Soc. Am. (submitted for publication).