Correlation of Speech Intelligibility Tests in Reverberant Rooms with Three Predictive Algorithms*

KENNETH D. JACOB

Bose Corporation, Framingham, MA 01701, USA

Correlation of Speech Intelligibility Tests in Reverberant Rooms with Three Predictive Algorithms*

KENNETH D. JACOB

Bose Corporation, Framingham, MA 01701, USA

The data base used in an earlier speech intelligibility study has been extended. The rooms added in the present study involve higher reverberation times and represent more difficult environments for sound reinforcement. The results of subjective intelligibility tests are compared to the predictions of three published algorithms. One of these predictors, based on the modulation transfer function, is being considered as an international standard (IEC Report 268-16, 1988), and its accuracy is confirmed in this study. A second predictor, based on the Lochner–Burger signal-to-noise ratio, is shown to be equally accurate when data are analyzed in a manner similar to that of two other intelligibility studies. The third algorithm, which assumes that all reflections degrade intelligibility, is shown to be inferior. The study confirms the beneficial effect of early reflections and the detrimental effect of late reflections on speech intelligibility assumed by the two more accurate algorithms.

0 INTRODUCTION

The clarity with which a sound system transmits speech is one of the fundamental parameters governing its quality. Acceptable speech intelligibility is therefore a basic objective of sound system designers. It is well known that frequency response, distortion, and background noise affect intelligibility, and the relationship between these factors and intelligibility has been studied extensively [1]–[3]. As a result, techniques have been developed to aid designers in predicting their effects. Another parameter known to affect speech intelligibility is reverberation, and studies to quantify its effect have also led to predictive algorithms. Two of these algorithms assume that early reflections contribute to intelligibility. A third assumes that all reflections degrade intelligibility and that intelligibility is monotonically related to loudspeaker directivity. Because of their different assumptions, it can be expected that these algorithms will produce different results when applied to reverberant environments.

In a preliminary study by this author [4], an experimental data base consisting of five rooms, three different loudspeakers, and two listener positions was used to test the accuracy of these same three predictive algorithms. The rooms were moderately reverberant, and the conclusions of the first study therefore were limited in scope. Because it is also important that sound system designers be capable of accurately predicting intelligibility in more reverberant environments, five rooms of higher reverberation time were added to the original data base. Conclusions reached in the present study can be applied to the broad range of reverberant environments in which sound systems typically operate.

In the present study, for each combination of room, loudspeaker type, and listener position, subjective intelligibility tests have been conducted to obtain estimates of the actual intelligibility. Against these actual intelligibility scores, the accuracy of each of the three predictive algorithms has been tested. The results are presented in side-by-side comparisons. In addition, the experiment has been documented to facilitate reproduction by other investigators. Methodology, experimental technique, and intermediate results are described explicitly.

1 SOURCE, ROOM, AND RECEIVER PARAMETERS

1.1 Loudspeakers

The three different loudspeaker sources used are listed in Table 1.

1.2 Room Parameters

A total of 10 rooms located in the Boston metropolitan area were studied, as given in Table 2.

* Presented at the 85th Convention of the Audio Engineering Society, Los Angeles, 1988 November 3–6; revised 1989 July 24 and September 25.

1.3 Source Positions

Sound sources were positioned in the part of the room normally used for speaking. In the rooms that had proscenium-type stage areas, loudspeakers were always placed in the middle and on the audience side of the proscenium. In some rooms, the loudspeakers were raised to approximately the middle top of the proscenium opening using a pneumatic tower, and in others, loudspeakers were placed upon 6-ft (2-m) stands, as shown in Fig. 1. In all cases, efforts were made to place loudspeakers in locations suitable for either permanent or temporary sound systems.

1.4 Listener Positions

In each room, two listener locations were selected, one corresponding approximately to one-third the distance from the loudspeaker to the rear wall, and the other at the rear of the auditorium. These positions were chosen in the expectation of yielding a wide range of intelligibility scores. In all cases, the relationship between loudspeakers and listeners was such that listeners were within 7.5° of the major axes of the loudspeakers.

2 SOUND-SYSTEM ADJUSTMENTS

To study the effect of reverberation on intelligibility, efforts were made to minimize the effects of other factors known to affect intelligibility. The loudspeakers were adjusted individually to have equivalent frequency response to within ±2 dB as measured on a one-third-octave analyzer. Bandwidth was limited to from 100 Hz to 10 kHz. Any audible noise, hum, or distortion from the loudspeakers was eliminated before tests were conducted.

Background noise was minimized as a factor by ensuring a ratio of speech level to background noise of at least 30 dB as measured on a instrumentation-grade sound pressure level meter set on A-weighting. In general, a ratio of speech signal to background noise of at least 15 dB is required to eliminate noise as a factor [2].

3 SUBJECTIVE TESTING

To establish an accurate estimate of the true speech intelligibility, subjective tests were conducted for every combination of room, sound source, and listener location. The subjective tests were administered according to ANSI S3.2-1960(R1982) [5].

In the ANSI test, subjects are presented with a combination of 20 lists of 50 phonetically balanced (PB) monosyllabic English words. When presented to the listener, the words are embedded in a carrier phrase, for example, "you will write the word *test* now." The carrier phrase is necessary to simulate the interfering effect of continuously spoken words. Words are scored according to phonetics rather than spelling; for example, "payed" is equivalent to "paid." If any part of the word is phonetically incorrect, the entire word is scored incorrect. The percent intelligibility is defined as the percentage of phonetically correct words, and is denoted by $\%PB_{act}$.

Word lists reproduced through the loudspeakers were originally recorded in an anechoic chamber using an omnidirectional instrumentation-grade microphone at a distance of 0.5 m from the talker in order to create an on-axis speech recording containing negligible reverberation.

Subjects were chosen from the general public. At least 3 years of college and normal hearing as checked by an audiometer were required. Subjects were not told the nature of the experiment.

In the first five rooms (see Table 2) subjective testing was conducted in the rooms. For the second group of five rooms, a Sennheiser dummy head and dual microphone combination was used at the two listener positions to record the word lists.[1] These binaural recordings were played back over headphones to subjects at a later date. Thus testing was binaural in all cases.

Testing was conducted over two periods of five successive days each. Different subjects were used for the two periods. Word lists were scrambled to avoid any long-term training of the listeners, and the same word lists were never played on the same sources or in the

[1] A preliminary experiment was conducted to quantify the accuracy of the binaural recording and playback system. Results showed negligible change in intelligibility scores when they were compared to scores from the same test taken with subjects in the room.

Table 1. Loudspeakers

Name	Type	Axial Directivity
Soundsphere 2212-1	Omni radiator	1.1
Bose 802-II	Eight-driver array	7.3
Electro Voice HR6040A (with TL806AX)	Constant-directivity horn	17.7

Table 2. Room parameters.

Name	T_{60} (s)	Volume (m^3)
Berklee Performance Center, Boston	0.9	5 450
Coolidge Corner Movie House, Brookline	1.0	4 590
Huntington Theater, Boston	1.1	3 190
Saint Bridget's Church, Framingham	2.0	3 810
Nevins Hall, Framingham	3.5	10 620
Jordan Hall, Boston	2.2	4 530
Mechanics Hall, Worcester	2.2	11 582
South End Cathedral, Boston	3.3	59 152
The Cyclorama, Boston	3.5	11 610
MIT Indoor Track, Cambridge	4.6	42 475

same listener positions from day to day. In addition, two lists were never played in succession for any one condition. The total number of words presented for each room, source, and listener position combination ranged from 2000 to 2800 words.

4 OBJECTIVE MEASUREMENTS

For each room, source, and listener position combination, system impulse responses representing the transfer functions from loudspeaker input terminals to listener were recorded. A microcomputer-based data-acquisition system was used to measure, store, and analyze the impulses. The sampling rate was 9615 Hz, representing a usable bandwidth of approximately 4 kHz. 8192-point buffers were used for each impulse.

The directivity of the constant-directivity horn was taken from the manufacturer's data sheet. For the omni radiator and the array loudspeaker, polar responses were taken in an anechoic chamber using multiple microphone locations in order to measure full-sphere radiation. The directivity of both of these devices is not constant in the frequency range above 1000 Hz due to the interaction of multiple drivers and diffraction effects. The directivities quoted in Table 1 represent averages of the 1-, 2-, and 4-kHz octave bands.

Reverberation time was calculated in each room as follows. For each impulse response, 1) the impulse was squared, 2) the squared impulse was integrated from maximum time (about 800 ms) to zero time (so-called reverse integration), 3) the resulting curve was converted to decibels to form the curve that would be measured if the room were filled with steady-state noise and then shut off, 4) a linear regression was applied to the curve from 100 to 600 ms after the onset of decay, 5) the reverberation time for the impulse was estimated as 60 divided by the decay rate of the regression line. Finally, the times for several impulses were averaged. However, reverberation times calculated for impulses within a given room were consistent within ±3%.

The room volume was calculated by analyzing blueprints. The distance from source to listener was computed from the impulse response by multiplying the delay time from source to measurement microphone by the speed of sound.

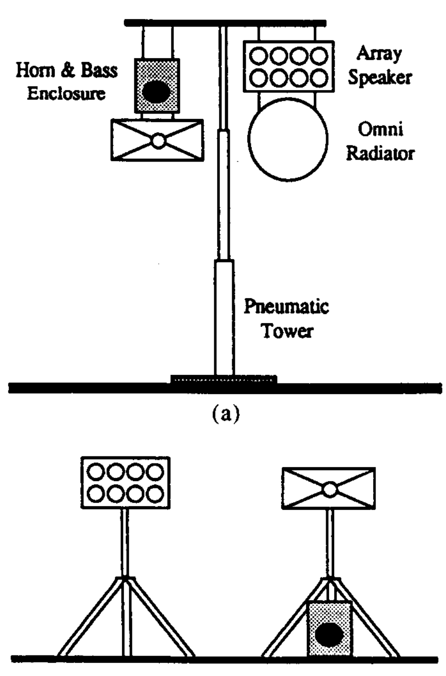

Fig. 1. Two different loudspeaker configurations used.

5 PREDICTIVE ALGORITHMS

5.1 Modulation Transfer Function (STI)

The modulation transfer function can be used to quantify the amount of degradation an input signal undergoes as it passes through a system [6]. The modulation method has been used as a quantitative measure of how organisms process visual signals, and in general for quantification of the performance of optical systems [7]. In room acoustics, signal degradation occurs in the time domain due to a variety of factors, including reverberation, distortion, and background noise. Because these factors are also known to degrade speech intelligibility, the modulation method was proposed by Houtgast and Steeneken as early as 1971 [8] as a quantitative predictor of intelligibility.

The algorithm for applying the modulation transfer function to the problem of predicting speech intelligibility uses a range of modulating frequencies (0.63–12.5 Hz) corresponding to the modulating frequencies of speech, and broadband noise corresponding to the average spectrum of speech. The modulation transfer function is calculated at discrete frequencies, weighted, summed, and normalized to yield a single index of speech intelligibility, called the speech transmission index (STI). Steeneken and Houtgast [9] showed that the technique was useful for a wide variety of different conditions, including reverberation, distortion, and background noise. Subsequently, the technique was simplified for the purpose of computational efficiency and is called the rapid speech transmission index (RASTI) [6].

Schroeder [10] derived the relationship between the impulse response of a system and its modulation transfer function. Schroeder's formula makes it practical to implement the modulation method on microcomputer-based data acquisition and analysis systems. This formula has been implemented in the present study:

$$m(F) = \left| \frac{\int_0^\infty p^2(t)\, e^{-j2\pi Ft}\, dt}{\int_0^\infty p^2(t)\, dt} \right| \qquad (1)$$

where

$m(F)$ = modulation transfer function
$p(t)$ = system impulse response
F = modulation frequency, hertz.

From Eq. (1) it can be seen that the modulation transfer function is proportional to the magnitude of the Fourier transform of the squared impulse response. More simply, the modulation transfer function is the very low frequency response of the squared impulse response. [It should be noted that direct implementation of Eq. (1) does not properly account for the effect of background noise on the STI. The STI algorithm specifies an input signal spectrum which is not flat, but rather approximates the average power spectrum of the human voice. Thus the signal-to-background-noise ratio (a factor directly affecting the STI) would be different than the ratio found by simply measuring the system's impulse response. In this experiment, however, a speech signal-to-background-noise ratio of more than 30 dB was ensured in all cases in order to minimize background noise as a factor. This allows Eq. (1) to be used with minimal error.]

From the fact that the STI is calculated for modulation frequencies $\leqslant$ 12.5 Hz, it can be shown that early reflections arriving within a certain time limit can never lower the STI; in fact, in the presence of late reverberation, early reflections increase the STI. Houtgast et al. [11] calculated the STI in a range of exponentially decaying systems. They used a variety of hypothetical cutoff times between useful and detrimental reflections and found that a cutoff of approximately 75 ms corresponded best to actual intelligibility. This agrees in general with the work of Lochner and Burger [12]. It is central to the modulation method that early reflections have a beneficial, and at worst no detrimental, effect on intelligibility.

5.2 Lochner–Burger Signal-to-Noise Ratio

Lochner and Burger [12] performed fundamental research on how speech energy, particularly in the form of room reflections, is processed by the hearing system in order to understand speech. They found that a percentage of the energy in a reflection is integrated, or summed, with the direct sound depending on both its level and the arrival time. For example, a reflection whose level is 5 dB less than the first arrival will be totally integrated by the ear if it arrives within 40 ms of delay, according to their findings.

Lochner and Burger showed that the integrated energy from early reflections could be lumped into a signal part, and energy from late arriving reflections combined with background noise lumped into a noise part, to form a signal-to-noise ratio for speech intelligibility. Several other investigators have used the same basic idea of a psychoacoustic signal-to-noise ratio to form objective measures of subjective impressions [13]. Lochner and Burger's formula for the signal-to-noise ratio is

$$R_e = 10 \log \frac{\int_0^{95\ \mathrm{ms}} p^2(t)\alpha(t, l)\ dt}{\int_{95\ \mathrm{ms}}^{\infty} p^2(t)\ dt} \qquad (2)$$

where

R_e = Lochner–Burger signal-to-noise ratio, decibels
$p(t)$ = impulse response of system
$\alpha(t, l)$ = weighting curve applied to reflections according to delay t and level l.

Lochner and Burger did not provide conclusive experimental data showing the exact relationship between their signal-to-noise ratio and speech intelligibility. Other investigators, most notably Bradley [14] and Latham [15], have found that the Lochner–Burger signal-to-noise ratio contains the information necessary to predict speech intelligibility. However, they found it necessary to develop their own correlation functions between objective and subjective measures.

5.3 %AL$_{cons}$ Formula for Directional Sources

Peutz [16] developed a subjective intelligibility test method which scored word lists according to the mean percentage of misunderstood consonants. The resulting score, the percent articulation loss of consonants (denoted here by $\%AL_{cons\ act}$) can be shown (see Sec. 6.2) to be simply related to $\%PB_{act}$ as measured and scored using the ANSI standard [5].

Peutz conducted an experiment using an omnidirectional loudspeaker in a variety of reverberant rooms. His data indicated that $\%AL_{cons}$ as measured using the subjective test increased linearly as a function of the square of the source-to-listener distance up to a certain limiting distance, beyond which there was no further increase. The data also indicated that this limiting distance was simply related to the critical distance of the room–loudspeaker combination, defined as the source-to-listener distance at which the direct field and the steady-state reverberant field are equal in intensity [17]. From these trends, a simple formula was written to predict $\%AL_{cons}$:

$$\%AL_{cons\ pred} = \begin{cases} 200D^2T^2/V, & D < D_L \approx 3.2D_C \\ 9T, & D > D_L \end{cases} \qquad (3)$$

where

D = source-to-listener distance, meters
T = reverberation time, seconds
V = room volume, cubic meters
D_L = limiting distance, meters
D_C = is critical distance, meters

No statistical analysis of the data was offered to show the accuracy with which the articulation loss could be predicted, except for the stated accuracy of ± 10%. A 10% error in $\%AL_{cons}$ corresponds to a 17% error in %PB as measured by the ANSI standard (see Fig. 3).

Klein [18] subsequently hypothesized that because the critical distance can be shown to be theoretically related to loudspeaker directivity through the use of a statistical reverberation formula, Peutz's limiting distance D_L could be extended by increasing directivity.

The statistical reverberation formula is

$$D_C = K\left(\frac{Q_{\mathrm{AXIAL}} V}{T}\right)^{1/2} \tag{4}$$

where

K = a constant
Q_{AXIAL} = source directivity on the axis from source to listener.

The modified Peutz formula introducing loudspeaker directivity Q was published by Klein [18], but is attributed to Davis [19],

$$\%\mathrm{AL}_{\mathrm{cons\ pred}} = \begin{cases} 200D^2T^2/Q_{\mathrm{AXIAL}}V, & D < D_L \approx 3.2D_C \\ 9T, & D > D_L. \end{cases} \tag{5}$$

No data were offered by Klein to substantiate his hypothesis. To this author's knowledge, no attempt has been made to correlate Eq. (5) with actual speech intelligibility scores except for this author's previous study, which indicated poor correlation [4]. This is significant in light of the fact that the equation has been implemented in many manuals, design guides, and computer programs for designing sound systems.

It should be stressed that $\%\mathrm{AL}_{\mathrm{cons\ act}}$ is a parameter specified by a particular subjective intelligibility test, analogous to the ANSI test. The $\%\mathrm{AL}_{\mathrm{cons\ act}}$ parameter must be separated from any formulas that have been written to predict $\%\mathrm{AL}_{\mathrm{cons}}$. In this study, one of these formulas, Eq. (5), has been chosen for examination. However, the subjective intelligibility test yielding $\%\mathrm{AL}_{\mathrm{cons\ act}}$ values is not under examination.

6 RESULTS

6.1 Results of Subjective Testing

Intelligibility was scored for each subject in each room, source, and listener position combination, and a mean value was computed. The resulting average is denoted by $\%\mathrm{PB}_{\mathrm{act}}$. The accuracy of these mean values was estimated to within a 95% confidence interval using Student's t test [20]. Results of the subjective testing are shown in the Appendix, Table A.1.

6.2 Converting Objective Data to $\%\mathrm{PB}_{\mathrm{pred}}$ Scores

The aim of this study is to compare the accuracy of three speech intelligibility predictors. The modulation and the signal-to-noise methods both predict the same parameter as used in this study's subjective testing, %PB, allowing a direct comparison of predicted versus actual scores.

STI values were converted to $\%\mathrm{PB}_{\mathrm{pred}}$ scores using the best-fit third-order regression curve established by Steeneken and Houtgast [9] for a wide variety of speech transmission conditions. The formula for this regression curve is

$$\%\mathrm{PB}_{\mathrm{pred}} = -43.0 + 279.8\mathrm{STI} - 31.2\mathrm{STI}^2 - 124\mathrm{STI}^3 . \tag{6}$$

The technique used to convert the Lochner–Burger signal-to-noise ratios to $\%\mathrm{PB}_{\mathrm{pred}}$ for each room, source, and listener position combination is similar to that used by Bradley [14] and Latham [15]. A second-order polynomial regression curve was calculated using $\%\mathrm{PB}_{\mathrm{act}}$ versus signal-to-noise ratio data, as shown in Fig. 2. The equation for the regression curve can then be used to convert signal-to-noise ratios to $\%\mathrm{PB}_{\mathrm{pred}}$ scores.

The regression line of Fig. 2 is represented by

$$\%\mathrm{PB}_{\mathrm{pred}} = 79.458 + 2.99R_e - 0.1368R_e^2 . \tag{7}$$

The percent articulation loss of consonants $\%\mathrm{AL}_{\mathrm{cons}}$ is a method of subjectively evaluating speech articulation which differs from methods used to obtain %PB scores. Similarly, a $\%\mathrm{AL}_{\mathrm{cons}}$ value predicted by a formula cannot be directly compared to a predicted %PB value. A relationship is required to convert $\%\mathrm{AL}_{\mathrm{cons}}$ scores to %PB scores.

To establish the relationship between $\%\mathrm{AL}_{\mathrm{cons}}$ and %PB, the word lists collected in this study were rescored according to the $\%\mathrm{AL}_{\mathrm{cons}}$ procedure as outlined by Peutz [16]. An excellent relationship was established between the two methods of scoring, as shown in Fig. 3, and

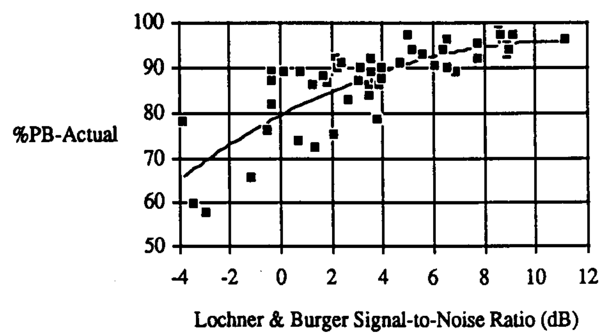

Fig. 2. Lochner–Burger signal-to-noise ratios versus $\%\mathrm{PB}_{\mathrm{act}}$ and regression line.

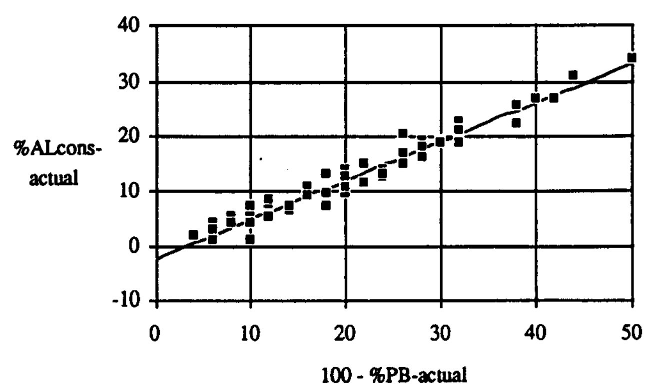

Fig. 3. Scattergram showing relationship between $\%\mathrm{AL}_{\mathrm{cons\ act}}$ and $\%\mathrm{PB}_{\mathrm{act}}$, both scored from the same subjective tests.

is used to convert $\%AL_{cons\ pred}$ to $\%PB_{pred}$ scores.

The regression line of Fig. 3 is represented by

$$\%PB_{act} = 100 - \frac{\%AL_{cons\ act} + 2.12}{0.70} . \quad (8)$$

To summarize, the STI scores were converted to $\%PB_{pred}$ scores using a previously published formula; Lochner–Burger signal-to-noise ratios were converted by calculating a best curve fit to the data collected in this study; and the predicted $\%AL_{cons}$ scores were converted by establishing a relationship between the %PB and $\%AL_{cons}$ methods of scoring intelligibility word lists. Using these conversions, the three predictive algorithms can be directly compared on $\%PB_{act}$ versus $\%PB_{pred}$ graphs.

6.3 Correlation of Predicted Scores with Subjective Scores

Scatter plots of $\%PB_{act}$ versus $\%PB_{pred}$ for each of the three predictive methods are shown in Fig. 4. The graphs can be interpreted visually by realizing that perfect predictions fall on the straight line. Points below the straight line represent predicted scores that are too high when compared to actual scores, and points above the line represent predicted scores that are too low.

The data show a relatively even distribution of points about the straight line in Fig. 4(a) and (b). Significantly more scattering and some bias toward points above the line can be seen in Fig. 4(c).

6.4 Overall Standard Deviations

The data presented in the scattergrams can be examined more closely, and the degree of accuracy of the predictor quantified. The most straightforward measure of overall accuracy is the standard deviation, which is formed by squaring and adding the errors between actual and predicted scores, dividing by the total number of points, and taking the square root of the result. The overall standard deviation of the three predictive methods is shown in Fig. 5. The chart shows that the signal-to-noise method has the lowest overall standard deviation, followed closely by the modulation method; the standard deviation for the $\%AL_{cons}$ method is significantly higher. It is interesting to note that the standard deviation of 7% found for the modulation method is almost exactly the same as that quoted by Steeneken and Houtgast [9]. Furthermore there is some agreement between the standard deviation of 11.5% found for the $\%AL_{cons}$ method compared to the 17% accuracy ($10\%AL_{cons}$) quoted by Peutz [16]. Different data bases, however, do not make these overall deviations exactly comparable.

6.5 Mean Errors

Another way to examine the data is to compute the mean differences between predicted and actual scores. This analysis yields information about whether the data contain any biases. For example, if most of the points in a scattergram fall below the straight line, meaning that the predictions were consistently too high, the mean difference would be positive. Zero mean difference signifies no biases in the predictor. The overall mean differences for the three predictors are shown in Fig.

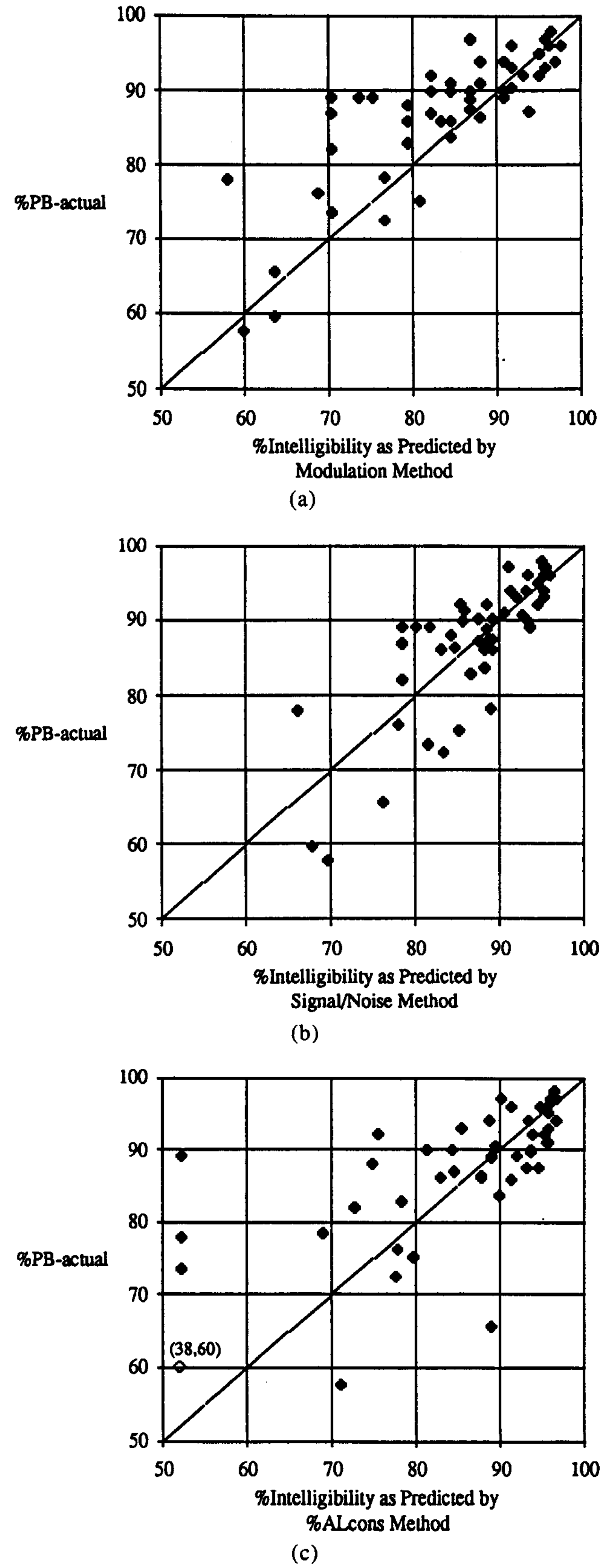

Fig. 4. Scattergrams of $\%PB_{act}$ versus $\%PB_{pred}$ for three predictive algorithms. (a) Modulation method. (b) Signal-to-noise method. (c) $\%AL_{cons}$ method.

6. The analysis shows no significant bias in the signal-to-noise procedure, an expected outcome of the second-order curve fitting (see Fig. 2). Both the modulation and the %AL_{cons} methods show a bias toward predictions, which are too low compared to the actual scores.

It must be stressed that the method used to convert Lochner–Burger signal-to-noise ratios to %PB_{pred} scores was developed from the data in this study only, and cannot be generalized without further study. The regression analysis was required because no universal method has been established by previous studies.

6.6 Correlation of Scores in Areas of Most Interest to Designers

Another way to examine the data is to ask whether there are any regions of special interest to sound system designers. In the context of designing sound systems, there are at least four. What is the accuracy of the three predictors when

1) Actual intelligibility is less than 85%?
2) Actual intelligibility is greater than 85%?
3) Predicted intelligibility is less than 85%?
4) Predicted intelligibility is greater than 85%?

The value of 85%PB intelligibility has been chosen because it is generally accepted as the transition between fair-to-good and good-to-excellent intelligibility. The first question addresses the accuracy of the predictors in cases where it is already known that the intelligibility is unacceptable, the second where actual intelligibility is known to be good. The third question is relevant for cases where predictions show unacceptable intelligibility, and the fourth where predictions show acceptable intelligibility. The results are shown in Fig. 7.

Fig. 7 shows that in each of the four regions of special interest to the sound system designer, the modulation and signal-to-noise methods have lower standard errors than the %AL_{cons} method. The data contained in the charts also indicate that prediction is poorer overall in the region below 85% intelligibility.

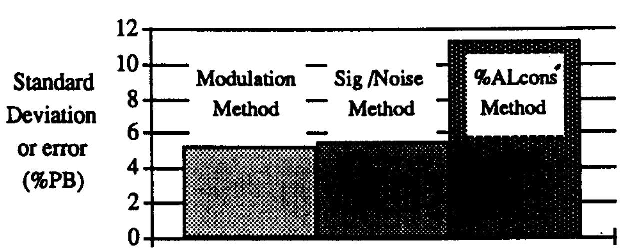

Fig. 5. Overall standard deviation of three predictors.

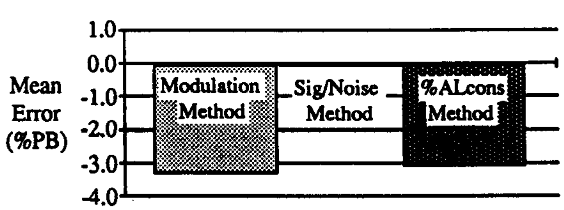

Fig. 6. Mean difference between predictive algorithms and actual intelligibility. Negative mean difference corresponds to predictions too low when compared to actual scores.

6.7 Illustrative Examples

Examination of the actual intelligibility data (see Appendix, Table A.1) reveals an unexpected case where the array loudspeaker was more intelligible than the horn loudspeaker of higher directivity (the rear position in the cathedral). In this case the modulation and signal-to-noise methods predict the same trend, while the %AL_{cons} method predicts that the horn should have been significantly more intelligible, as shown in Fig. 8.

Another unexpected result was observed when intelligibility was measured to be higher in the rear position than the front position; the horn in Mechanics Hall is an example. Again, the modulation and signal-to-noise methods predict the same trend, while the %AL_{cons} method predicts the opposite, as shown in Fig. 9.

These individual conditions suggest the need to in-

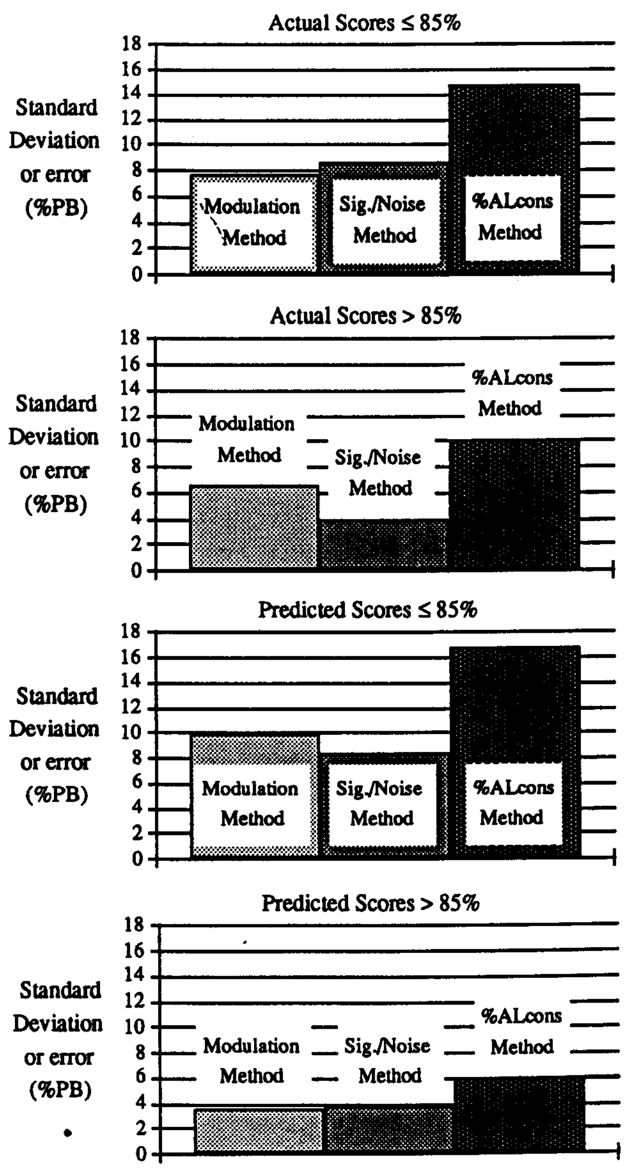

Fig. 7. Standard deviations for four special cases of interest to the sound system designer.

clude early reflections in the prediction of intelligibility. Furthermore, these cases show that while the additional direct sound provided by an increase in source directivity is useful for improving intelligibility, neglect of other factors affecting the ratio of early to late sound can lead to serious errors.

7 CONCLUSION

The data presented here support the conclusion that the speech transmission index, computed using the modulation method, is an accurate predictor of speech intelligibility in reverberant rooms. Correlation between actual and predicted intelligibility scores was very good overall, in both the under 85% and the over 85% regions. These results provide additional evidence of the suitability of the STI as an international standard for predicting speech intelligibility.

The Lochner–Burger signal-to-noise method was shown to be accurate as a predictor when a function unique to the data base of this study was used to translate signal-to-noise ratios to $\%PB_{pred}$ scores. Two other investigations using the method required different functions [14], [15] for this purpose. Thus while the signal-to-noise ratio has been shown to contain the information necessary to predict intelligibility within this data base accurately, no universal method of translating signal-to-noise ratios to %PB intelligibility has been established. Additional investigation is required to establish a general relationship between the Lochner–Burger signal-to-noise ratio and intelligibility.

The formula for predicting $\%AL_{cons}$, as defined in Eq. (5), has been shown to be the least accurate. From the results of this extended data base it must be concluded that intelligibility as predicted by the formula can lead to serious errors. In addition, the hypothesis that intelligibility can be increased monotonically simply by increasing loudspeaker directivity should be rejected. It is important to note that under the same speech conditions, the modulation and signal-to-noise methods predict intelligibility accurately, independent of loudspeaker directivity. The data therefore support the conclusion that loudspeaker directivity is important only in so far as it affects the ratio of early to late sound.

Finally, this experiment has been conducted and presented in such a manner as to allow repetition by other investigators. The hope is that the data base of rooms, sound sources, and listener positions can be

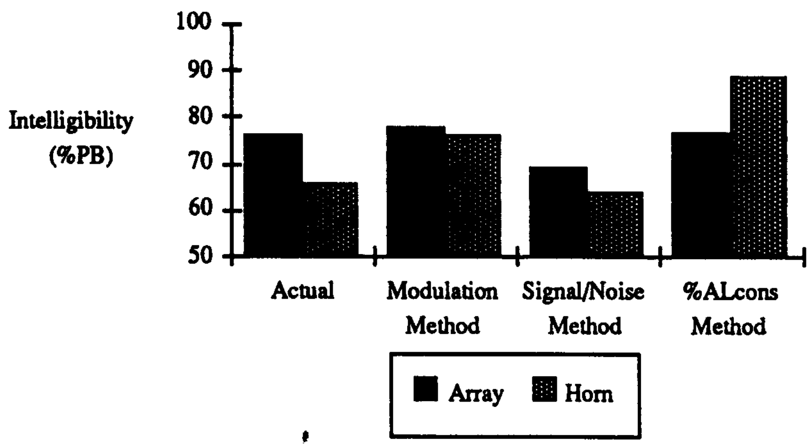

Fig. 8. Actual and predicted intelligibility scores in the rear of the South End Cathedral. Notice that the modulation and signal-to-noise methods predict the same trend as actual scores, while the $\%AL_{cons}$ method predicts the opposite trend.

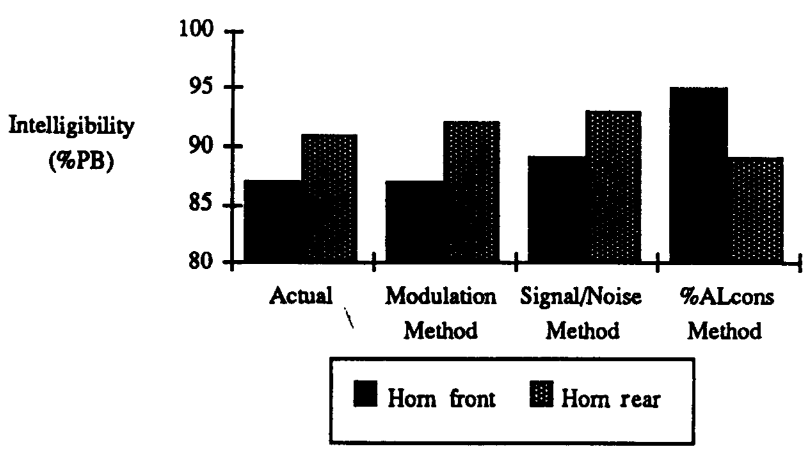

Fig. 9. Actual and predicted intelligibility scores in the front and rear of Mechanics Hall using the horn loudspeaker (see Appendix, Table A.1). Notice that the modulation and signal-to-noise methods predict the same trend as actual scores, while the $\%AL_{cons}$ method predicts the opposite.

further extended in an effort to establish increasingly more accurate predictors of intelligibility.

8 ACKNOWLEDGMENT

The author gratefully acknowledges the help of Dr. Amar Bose in the preparation of this manuscript, and Chris Ickler for discussions on the analysis and interpretation of data.

9 REFERENCES

[1] N. R. French and J. C. Steinberg, "Factors Governing the Intelligibility of Speech Sounds," *J. Acoust. Soc. Am.*, vol. 19, pp. 90–119 (1947).

[2] K. D. Kryter, "Methods for the Calculation and Use of the Articulation Index," *J. Acoust. Soc. Am.*, vol. 34, pp. 1689–1697 (1962).

[3] C. V. Paclovic, "Derivation of Primary Parameters and Procedures for Use in Speech Intelligibility Predictions," *J. Acoust. Soc. Am.*, vol. 82, no. 2 (1987).

[4] K. D. Jacob, "Subjective and Predictive Measures of Speech Intelligibility—The Role of Loudspeaker Directivity," *J. Audio Eng. Soc.*, vol. 33, pp. 950–955 (1985 Dec.).

[5] ANSI Std. S3.2-1960 (R1982), "American National Standard Method for Measurement of Monosyllabic Word Intelligibility," Am. Nat. Stand. Inst., New York.

[6] T. Houtgast and H. J. M. Steeneken, "A Review of the MTF Concept in Room Acoustics and Its Use for Estimating Speech Intelligibility in Auditoria," *J. Acoust. Soc. Am.*, vol. 77, no. 3 (1985).

[7] *Opt. Acta*, vol. 2, no. 18 (1971).

[8] T. Houtgast and H. J. M. Steeneken, "Evaluation of Speech Transmission Channels by Using Artificial Signals," *Acustica*, vol. 25 (1971).

[9] H. J. M. Steeneken and T. Houtgast, "A Physical Method for Measuring Speech Transmission Quality," *J. Acoust. Soc. Am.*, vol. 67, no. 1 (1980).

[10] M. R. Schroeder, "Modulation Transfer Functions: Definition and Measurement," *Acustica*, vol. 49 (1981).

[11] T. Houtgast, H. J. M. Steeneken, and R. Plomp, "Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function; Part 1: General Room Acoustics," *Acustica*, vol. 46, no. 1 (1971).

[12] J. P. A. Lochner and J. F. Burger, "The Influence of Reflections on Auditorium Acoustics," *J. Sound Vib.*, vol. 1, no. 4 (1964).

[13] H. Kuttruff, *Room Acoustics*, 2nd ed. (Appl. Sci. Publ., Ltd., London, 1979), p. 180.

[14] J. S. Bradley, "Predictors of Speech Intelligibility in Rooms," *J. Acoust. Soc. Am.*, vol. 80, no. 3 (1986).

[15] H. G. Latham, "The Signal-to-Noise Ratio for Speech Intelligibility—An Auditorium Acoustics Design Index," *Appl. Acoust.*, vol. 12, pp. 253–320 (1979).

[16] V. M. A. Peutz, "Articulation Loss of Consonants as a Criterion for Speech Transmission in a Room," *J. Audio Eng. Soc.*, vol. 19, pp. 915–919 (1971 Dec.).

[17] L. L. Beranek, *Acoustics* (Am. Inst. of Physics, New York, 1986).

[18] W. Klein, "Articulation Loss of Consonants as a Basis for the Design and Judgment of Sound Reinforcement Systems," *J. Audio Eng. Soc.*, vol. 19, pp. 920–922 (1971 Dec.).

[19] D. Davis, private communication, 1985.

[20] See, for example, H. Kohler, *Statistics for Business and Economics* (Scott, Foresman and Co., Glenview, IL, 1985).

APPENDIX

Table A.1. Subjective scores.*

Room	Source	Position	Score
Berklee	Sphere	Near	96 ± 1.0
		Far	93 ± 1.0
	Array	Near	96 ± 1.0
		Far	96 ± 1.0
	Horn	Near	98 ± 0.7
		Far	96 ± 1.0
Coolidge	Sphere	Near	97 ± 0.6
		Far	90 ± 1.6
	Array	Near	97 ± 0.6
		Far	94 ± 1.1
	Horn	Near	97 ± 0.6
		Far	91 ± 1.2
Huntington	Sphere	Near	94 ± 1.3
		Far	86 ± 1.7
	Array	Near	95 ± 1.0
		Far	89 ± 1.4
	Horn	Near	94 ± 1.1
		Far	92 ± 1.1
Bridget's	Sphere	Near	92 ± 1.1
		Far	82 ± 2.1
	Array	Near	92 ± 1.2
		Far	88 ± 1.5
	Horn	Near	93 ± 1.2
		Far	86 ± 1.5
Nevins	Sphere	Near	78 ± 3.0
		Far	89 ± 1.6
	Array	Near	87 ± 3.0
		Far	89 ± 1.1
	Horn	Near	89 ± 1.6
		Far	90 ± 1.6
Jordan	Array	Near	89 ± 1.4
		Far	78 ± 2.4
	Horn	Near	90 ± 1.9
		Far	87 ± 1.6
Mechanic's	Array	Near	86 ± 2.0
		Far	83 ± 2.0
	Horn	Near	87 ± 2.0
		Far	91 ± 1.6
Cathedral	Array	Near	90 ± 1.8
		Far	76 ± 3.7
	Horn	Near	91 ± 1.7
		Far	66 ± 2.6
Cyclorama	Array	Near	86 ± 2.6
		Far	73 ± 2.3
	Horn	Near	87 ± 2.3
		Far	72 ± 2.9
MIT Track	Array	Near	75 ± 2.5
		Far	60 ± 3.2
	Horn	Near	84 ± 2.9
		Far	58 ± 3.4

* Room—auditoriums listed in Table 2; Source—loudspeakers given in Table 1; Position—listener positions defined in Section 1.4; Score—mean over all subjects with the accuracy relative to the true intelligibility computed using Student's *t* test.

Table A.2. Speech transmission indexes, signal-to-noise ratios, and predicted $\%AL_{cons}$ for all conditions.*

Room	Source	Position	STI	L/B ratio	$\%AL_{cons\ pred}$
Berklee	Sphere	Near	0.65	6.6	8
		Far	0.71	8.9	8
	Array	Near	0.72	8.8	1
		Far	0.72	8.9	5
	Horn	Near	0.73	8.6	1
		Far	0.78	11.2	2
Coolidge	Sphere	Near	0.60	5.0	5
		Far	0.56	3.2	9
	Array	Near	0.71	9.2	1
		Far	0.64	6.4	3
	Horn	Near	0.71	8.7	0
		Far	0.61	4.7	1
Huntington	Sphere	Near	0.61	5.2	6
		Far	0.57	3.5	10
	Array	Near	0.70	7.8	1
		Far	0.64	6.9	4
	Horn	Near	0.74	9.0	0
		Far	0.67	7.8	1
Bridget's	Sphere	Near	0.56	2.2	17
		Far	0.48	−0.3	17
	Array	Near	0.70	3.6	3
		Far	0.54	1.7	17
	Horn	Near	0.65	5.6	1
		Far	0.54	1.3	7
Nevins	Sphere	Near	0.41	−3.8	32
		Far	0.48	−0.3	32
	Array	Near	0.48	−0.3	11
		Far	0.51	0.8	32
	Horn	Near	0.50	0.2	4
		Far	0.60	4.0	12
Jordan	Array	Near	0.60	3.6	7
		Far	0.52	3.8	20
	Horn	Near	0.64	6.6	2
		Far	0.56	3.1	9
Mechanic's	Array	Near	0.58	3.9	5
		Far	0.54	2.7	16
	Horn	Near	0.60	4.0	2
		Far	0.65	6.1	6
Cathedral	Array	Near	0.58	2.3	3
		Far	0.47	−0.5	17
	Horn	Near	0.58	2.4	1
		Far	0.44	−1.1	6
Cyclorama	Array	Near	0.61	1.9	8
		Far	0.48	0.7	32
	Horn	Near	0.68	3.7	3
		Far	0.52	1.4	14
MIT Track	Array	Near	0.55	2.1	15
		Far	0.44	−3.4	41
	Horn	Near	0.58	3.5	5
		Far	0.42	−2.9	19

* STI—speech transmission index generated from modulation tranfer function, as described in Section 5.1; L/B Ratio—measured Lochner–Burger signal-to-noise ratio as discussed in Section 5.2; $\%AL_{cons\ pred}$—predicted percent articulation loss of consontants as discussed in Section 5.3.

THE AUTHOR

Ken Jacob is a staff engineer of Bose Corporation, Framingham, Massachusetts. He received his master's degree from the Massachusetts Institute of Technology and his bachelor's degree from the University of Minnesota, both with specializations in acoustics.

In addition to research in speech intelligibility, Mr. Jacob is the acoustic engineer of the Bose 102[1] commercial sound system and the Acoustic Wave[1] Cannon system, and is the project engineer of the Sound System[1] software family including the Modeler[1] design program.

---

[1]Trademarks of Bose Corporation.