Subjective and Predictive Measures of Speech Intelligibility—The Role of Loudspeaker Directivity*

KENNETH D. JACOB

Bose Corporation, Framingham, MA 01701, USA

An experiment has been conducted to determine whether the speech intelligibility in rooms is related in a simple way to the loudspeaker directivity Q. Three loudspeakers of widely differing Q were used to subjectively test intelligibility in five auditoria. These results indicate that intelligibility and Q are not directly related. In addition, impulse response measurements were made so that several methods of predicting intelligibility could be compared with subjective scores. One method, which assumes a linear relationship between Q and intelligibility, was shown to be the least accurate predictor. Two other methods, one based on the psychophysics of the auditory system and the other based on the modulation transfer function, proved to be better predictors of intelligibility.

0 INTRODUCTION

Installed sound systems for speech and music are judged primarily by the answers to four questions: Is the system loud enough? Is the sound coverage even? Is the frequency response acceptable? Is speech intelligible?

In many situations, an installed system is used exclusively, or at least primarily, for speech reinforcement and reproduction. In these cases system intelligibility becomes the primary criterion upon which a system's performance is judged.

Loudspeaker directivity Q is generally thought of as a parameter affecting speech intelligibility. Subjective experiments designed to determine this relationship have not, to this author's knowledge, been the subject of published research. Based only on theoretical considerations, a commonly used technique for predicting speech intelligibility (after Klein [1]) assumes a linear relationship between Q and intelligibility;

$$\%\mathrm{AL}_{\mathrm{cons}} = \frac{200D^2T^2}{QV} \qquad (1)$$

where

$\%\mathrm{AL}_{\mathrm{cons}}$ = percentage articulation loss of consonants
D = source-to-listener distance
T = reverberation time of room
V = room volume
Q = loudspeaker directivity

In order to obtain a better understanding of sound system performance, an experiment was designed to answer two questions: 1) Is there a clear relationship between loudspeaker directivity and speech intelligibility? 2) Which of several techniques for predicting speech intelligibility are most accurate?

These will include Klein's [1] formula [Eq. (1)]; Lochner and Burger's procedure [2], which is based on the degree to which the auditory system integrates room reflections with direct sound; and the modulation transfer function (MTF), which quantifies the blurring effect reverberation has on speech [3].

Five auditoria were chosen for their broad range of reverberation times (0.9–3.5 s) and intended applications (cinema, theater, and meeting hall). In each room, two listening locations were chosen, roughly in the middle and at the rear of the auditorium floor. Using 14 trained listeners and three loudspeakers having dif-

* Manuscript received 1985 May 27, revised 1985 September 25.

ferent Q values, physical and subjective measurements were made.

The subjective results indicate that there is no simple relationship between Q and intelligibility, as measured by actual intelligibility tests. Even in reverberant rooms, where it might be assumed that a very directional loudspeaker would be required, medium and high Q loudspeakers performed nearly identically. Analysis of physical data revealed the formula which assumes a linear relationship between Q and intelligibility [Eq. (1)] to be the least accurate of the three predictive techniques. Furthermore, the two other techniques are better at predicting intelligibility over a wider range of situations.

1 EXPERIMENTAL DESIGN

The goal of the experimental design was to investigate loudspeaker directivity and its effect on intelligibility. Other variables known to affect intelligibility were held constant.

Three loudspeakers of differing directivities were used[1]:

High Q (Q = 17)	constant-directivity horn
Medium Q (Q = 7.5)[2]	array of identical drivers
Low Q (Q = 1.0)	spherical source

Five auditoria with various acoustical qualities, including two rooms known for their intelligibility problems, were chosen. Room parameters relating to this study are presented in Table 1.

Subjects were chosen from the general public. Each was screened for normal hearing as defined by American National Standards Institute (ANSI) Std S3.2-1960 [4]. Qualified subjects were trained using the same standard's guidelines. In anticipation of testing in rooms with known intelligibility problems, additional training was conducted in which background noise was used intentionally in order to create difficult conditions for intelligibility. The same 14 subjects were used throughout the experiment.

Table 1. Room parameters.

Room	T (seconds)	V (m^3)
Berklee Performance Center (primarily music)	0.9	5 450
Coolidge Corner Cinema (cinema)	1.0	4 590
Huntington Theater (speech)	1.1	3 190
Saint Bridget's Church (primarily speech)	2.0	3 810
Nevins Hall (primarily speech)	3.5	10 620

Intelligibility test material was in the form of monosyllabic English words embedded in a carrier sentence. Twenty lists of 50 words each, as defined by the ANSI standard, were used. These lists have the attribute of being phonetically balanced—in other words, individual speech sounds are represented with about the same frequency as in normal speech. Furthermore, the lists are approximately equivalent in difficulty.

Word lists were read by two different talkers at a rate of about 15 words per minute. Recordings of the word lists were made in an anechoic chamber using an instrumentation-grade omnidirectional microphone and a talker-to-microphone distance of 0.5 m. Thus a spectrally accurate on-axis speech recording was made.

A portable pneumatic tower was used to configure and elevate the three loudspeakers. The medium and high Q loudspeakers were individually aimed to provide the best coverage over the listening positions. Fig. 1 shows a sketch of the loudspeakers and tower in a typical location. In each room this location could be described approximately as the middle top of the stage proscenium. The configuration of the three loudspeakers was held constant from day to day.

Special care was taken to eliminate any hum or distortion in the test loudspeaker systems. In addition, because it was necessary to eliminate differences in loudspeaker-to-loudspeaker frequency responses as a variable in this experiment, a one-third-octave real-time analyzer was used at the listener locations to equalize the energy responses of each of the three sources. This was accomplished to within ±2 dB. In all cases the subjective quality of the speech being reproduced was very similar.

Finally, in order to minimize background noise as a variable affecting speech intelligibility, a speech signal-to-noise ratio of greater than 30 dBA was maintained. This was ascertained by measuring the background noise and adjusting the speech system gain so as to guarantee the signal-to-noise ratio. This is the accepted signal-to-noise ratio beyond which background noise has no significant effect on intelligibility [5].

In each room, two listener locations were chosen to coincide roughly with 1) the critical distance of the

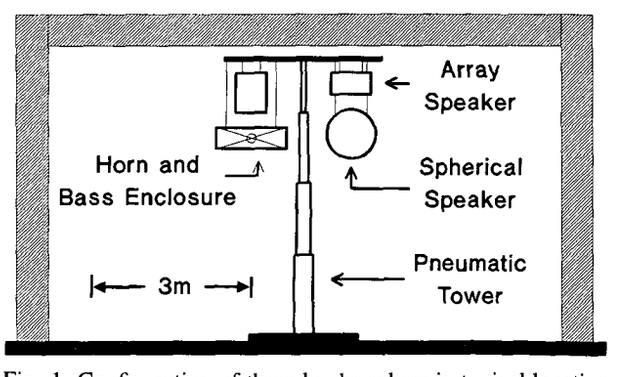


Fig. 1. Configuration of three loudspeakers in typical location in room.

[1] The terminology of high, medium, and low Q is used here to reflect the importance given Q in the $\%AL_{cons}$ formula. It is understood that higher Q sources exist.

[2] Array-type loudspeakers do not in general exhibit Q values as a function of frequency which are as stable as those of constant-directivity horns. The array loudspeaker used here has the following octave-band Q values: Q = 7.9 at 1 kHz, Q = 4.5 at 2 kHz, and Q = 9.5 at 4 kHz. The value given and used here for computations is an average of these three octave bands.

high Q source, and 2) the "intelligibility distance" of the high Q source (Fig. 2). Thus the listeners would would be within the distance range where intelligibility is predicted by Eq. (1) to vary directly with loudspeaker directivity.

The testing was carried out over five successive days. In each room, in each position, and with each source, four word lists were played. Or, in other words, there were 2800 data points for each loudspeaker in each position ($4 \times 50 \times 14 = 2800$). There was never more than one list in a row played successively over a given loudspeaker. In addition, the order of the 24 lists was changed from day to day. Logos were omitted from loudspeaker products, and subjects were not told the purpose or methodology of the experiment.

For each loudspeaker, in each position, and in each room, impulse response measurements were taken and stored digitally. (Once a system's impulse response has been captured, a system's frequency response, reflection arrival times, and reverberation time can be found.)

2 RESULT OF SUBJECTIVE TESTING

Word lists were scored by percentage of phonetically correct words. (Words could be spelled incorrectly and still be scored correctly, so long as they were phonetically correct, as specified by ANSI Std S3.2-1960 [4].)

In order to determine the accuracy of average subjective scores, statistical analysis was performed on the data using the Student's t test. Subjective scores are shown in Table 2. Using the t test and a confidence level of 95%, average subjective intelligibility scores were generally within an accuracy of 1–2%.

Fig. 3 is a bar chart of the average intelligibility scores. From Table 2 and Fig. 3 it can be seen that there is little or no statistical difference between the scores of the high and medium Q loudspeakers, while the low Q source is at times significantly less intelligible. These data indicate that the theory which assumes intelligibility to be directly proportional to loudspeaker directivity [Eq. (1)] may not be correct.

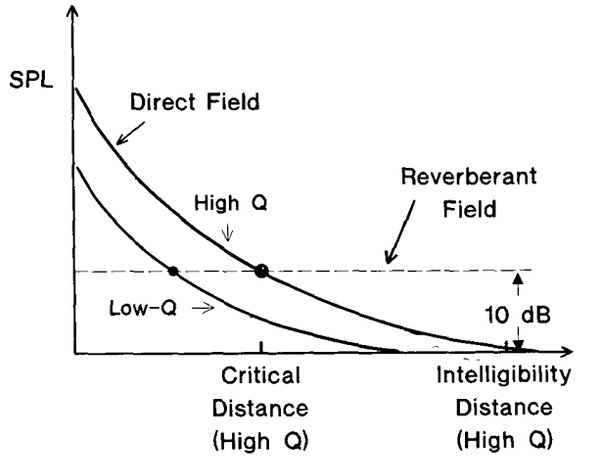


Fig. 2. Critical and "intelligibility" distances for two different Q loudspeakers as defined by classical statistical reverberation theory.

3 PREDICTIVE TECHNIQUES

Three well-known techniques are used for predicting speech intelligibility in rooms:

3.1 Articulation Loss of Consonants

Peutz [6] used an omnidirectional loudspeaker in a variety of rooms in order to investigate intelligibility empirically. He found that subjective intelligibility could be based on the percentage of correctly understood consonants in special monosyllabic nonsense words. Peutz showed that articulation loss varied with the square of the source-to-listener distance. This relationship held until a certain distance was reached, beyond which the articulation loss remained constant; he called this the critical distance. (In order to avoid confusion with the more traditional definition of critical distance, namely, the distance at which the direct and reverberant fields from a source in a room are equal, we shall call Peutz's critical distance the *intelligibility distance*.) The intelligibility distance is about 3.2 times greater than the critical distance. Peutz's formula for the articulation loss of consonants is

$$\%\mathrm{AL}_{\mathrm{cons}} = \frac{200D^2T^2}{V} \tag{2}$$

where

$\%\mathrm{AL}_{\mathrm{cons}}$ = percentage articulation loss of consonants
D = source-to-listener distance
T = reverberation time of room
V = room volume

Klein [1] modified Peutz's formula for articulation loss by including loudspeaker directivity. He utilized the fact that critical distance is related to loudspeaker directivity by

$$C_{\mathrm{d}} = \left[\frac{QV}{T}\right]^{1/2} \tag{3}$$

where

C_{d} = critical distance
Q = loudspeaker directivity

Table 2. Subjective intelligibility scores.

Room/ position	High Q	Medium Q	Low Q
1 1	98 ± 0.7	96 ± 1.0	96 ± 1.0
1 2	96 ± 1.0	96 ± 1.0	93 ± 1.0
2 1	97 ± 0.6	97 ± 0.6	97 ± 0.6
2 2	91 ± 1.2	94 ± 1.1	90 ± 1.6
3 1	94 ± 1.1	95 ± 1.0	94 ± 1.3
3 2	92 ± 1.1	89 ± 1.4	86 ± 1.7
4 1	93 ± 1.2	92 ± 1.2	92 ± 1.1
4 2	86 ± 1.5	88 ± 1.5	82 ± 2.1
5 1	89 ± 1.6	87 ± 3.0	78 ± 3.0
5 2	90 ± 1.6	89 ± 1.1	89 ± 1.6

Klein assumed that Peutz's formula could be modified to account for the dependence of the critical distance on loudspeaker directivity. The resulting formula is

$$\%AL_{cons} = \frac{200D^2T^2}{QV} .$$

3.2 Signal-to-Noise Procedure

Lochner and Burger [2] concentrated upon the fact that speech intelligibility was dependent on, among other things, the ratio of speech signal to background noise. In early work they established this relationship through subjective testing. They later hypothesized that this basic relationship between speech signal and masking noise could be adapted to include reverberation. They reasoned that for a certain period of time the hearing system integrates energy in the form of room reflections with the sound energy arriving directly from the source. (This is a well-known phenomenon in psychoacoustics [7].) They designed an experiment to determine the degree to which reverberant energy was integrated by the hearing system. They postulated that the portion of energy integrated could be considered signal and that the remaining reverberant energy could be considered noise. This led to a formula for the effective signal-to-noise ratio,

$$S/N_{eff} = \frac{\int_0^{95\text{ ms}} p^2(t)a(t)\ dt}{\int_{95\text{ ms}}^{\infty} p^2(t)\ dt} \quad (5)$$

where

S/N_{eff} = effective signal-to-noise ratio
$p(t)$ = impulse response of system
$a(t)$ = weighting function for integration properties of the hearing system

The effective signal-to-noise ratio was then used in conjunction with their subjective data to predict intelligibility.

Lochner and Burger made both subjective and physical measurements in a variety of rooms and found excellent agreement between predicted and measured values of speech intelligibility. Other investigators [8] have also found excellent correlation.

3.3 Modulation Transfer Function

The modulation transfer function (MTF) technique for predicting intelligibility relies on the fact that reverberation and background noise have the effect, at the output of a system, of smearing, or blurring, an input waveform. The modulation transfer function was a technique first used to measure the accuracy and clarity of optical systems. Houtgast and Steeneken [3] adapted the modulation transfer function in order to predict intelligibility in speech transmission channels.

In the modulation transfer function technique, speech-band noise (the carrier) is modulated by frequencies which coincide with the modulating frequencies of natural speech. As the modulated noise passes through a speech transmission system, the smearing effect can be measured by the change from input to output of the modulation depth.

Once the modulation transfer function has been generated, it is weighted and summed to yield a single number, the speech transmission index (STI), which is an indicator of speech intelligibility. Houtgast and Steeneken have found very good correlation between predicted and measured values using a variety of systems.

Schroeder [9] derived that connection between the impulse response of a system and its modulation transfer function,

$$m(F) = \left| \frac{\int_0^{\infty} p^2(t)\ e^{-j\omega t}\ dt}{\int_0^{\infty} p^2(t)\ dt} \right| \quad (6)$$

where

$p(t)$ = system impulse response
$m(F)$ = modulation transfer function

In words, the modulation transfer function is proportional to the magnitude of the Fourier transform of the squared impulse response.

4 ANALYSIS

Klein's formula for predicting speech intelligibility was calculated by computing room volume, reverberation time, source-to-listener distance, and source directivity Q. It should be noted that reverberation time was taken from the impulse response measurements and was not calculated using predictive methods, which can cause significant errors.

Lochner and Burger's signal-to-noise procedure was calculated by processing the digitally stored impulse responses according to Eq. (5). The modulation transfer function was generated according to Eq. (6).

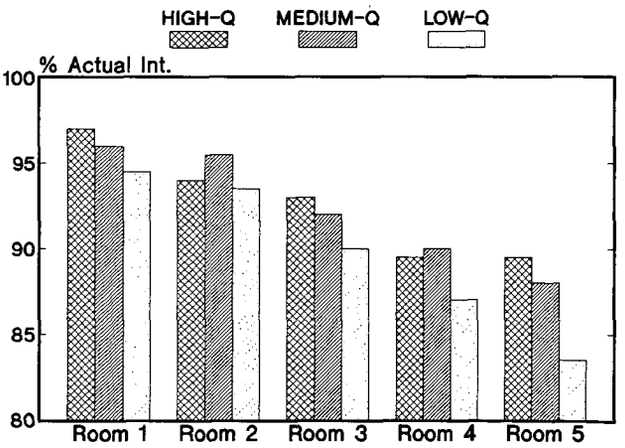


Fig. 3. Actual subjective intelligibility scores for three different Q loudspeakers in five rooms, averaged over two positions.

5 SUBJECTIVE VERSUS PREDICTIVE INTELLIGIBILITY SCORES

Results of the three predictive methods were translated using mapping graphs provided by Houtgast and Steeneken [3] and Lochner and Burger [2] into percentage phonetically balanced (PB) word intelligibility scores so that they could be compared with the subjective data.

Fig. 4 shows such a comparison for each of the three predictive methods in the five rooms. (If the predictive methods were ideal, all points would fall on the line indicated in Fig. 4.) These plots show immediately that the predictive method that assumes a linear relationship between Q and intelligibility [Eq. (1)] has the most scattering and is therefore the least accurate. These deviations can be examined more closely. Fig. 5 shows the variance of predicted versus actual values for each predictive technique in each room. In almost all cases, the variance produced by using Eq. (1) is greater than that due to the other two techniques.[3] It is also important to note that the variance in points predicted by Klein's formula is highest in rooms where the reverberation time is high (rooms 4 and 5). In comparison, the variance of the other two techniques also increases, but the levels are much lower. Another way to interpret the results is to compute the mean differences between predicted and actual intelligibility scores. Fig. 6 shows these differences. It is clear from these data that the signal-to-noise method consistently predicts intelligibility scores that are 2–5% too high, whereas the modulation transfer function method predicts scores that are 3–6% too low.

6 DISCUSSION

The data show that Klein's method of predicting intelligibility can be inaccurate, especially in highly reverberant rooms. In these cases the formula predicts intelligibility scores too high for the high Q source and too low for the low and medium Q loudspeakers. These are rooms where the prediction of intelligibility is most crucial since they are most likely to have intelligibility problems.

In general, as measured by actual intelligibility tests, high and medium Q loudspeakers performed equally well. The low Q loudspeaker performed significantly less well in some cases, especially in the two reverberant rooms. These results can be explained in an intuitive way by considering the impulse responses typical of the three systems.

The high Q source provides the highest ratio of direct to reverberant energy. The medium Q device has a lower ratio of direct to reverberant energy, but provides, in most rooms, much more energy in the form of early reflections than the high Q source. So long as these early reflections are early enough to be integrated by the hearing system, they can be considered signal, and therefore contribute to intelligibility by improving the

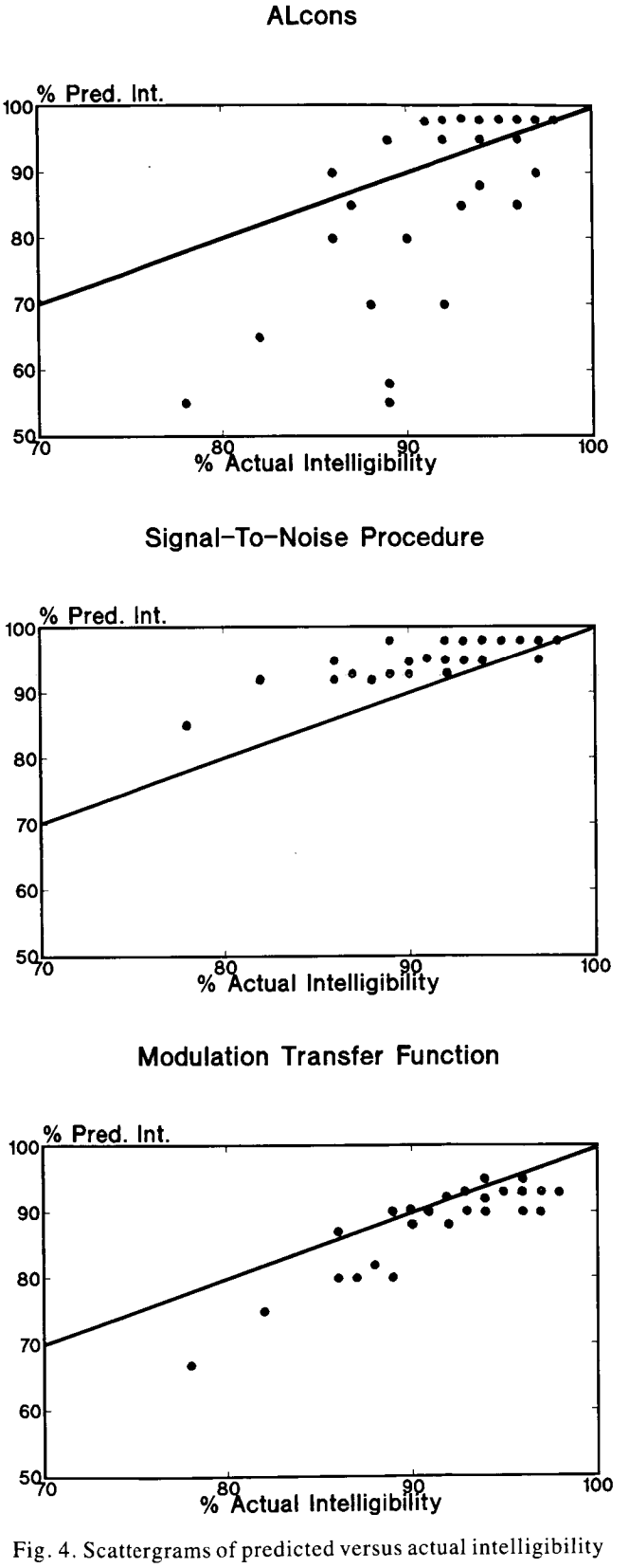


Fig. 4. Scattergrams of predicted versus actual intelligibility scores.

[3] Houtgast and Steeneken [3] have shown that for the theoretical case of a room whose impulse response is perfectly exponential, speech intelligibility as predicted by the modulation transfer function and Eq. (1) will be the same. However, real rooms usually deviate substantially from this ideal. Intelligibility as predicted by the modulation transfer function method takes these deviations into account while the $\%AL_{cons}$ method Eq. (1) does not; thus predicted scores can be significantly different.

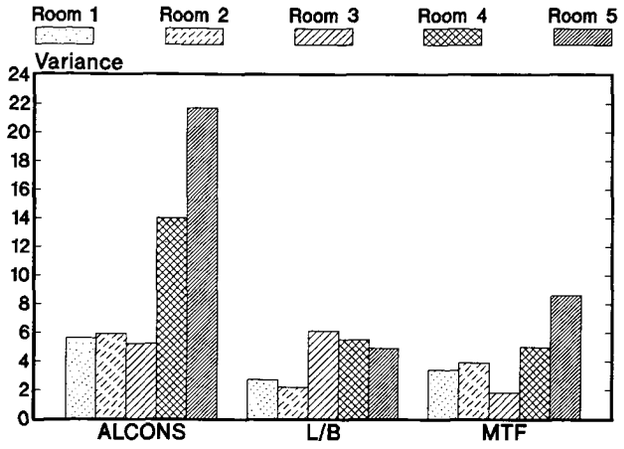


Fig. 5. Variance of predicted versus actual intelligibility scores.

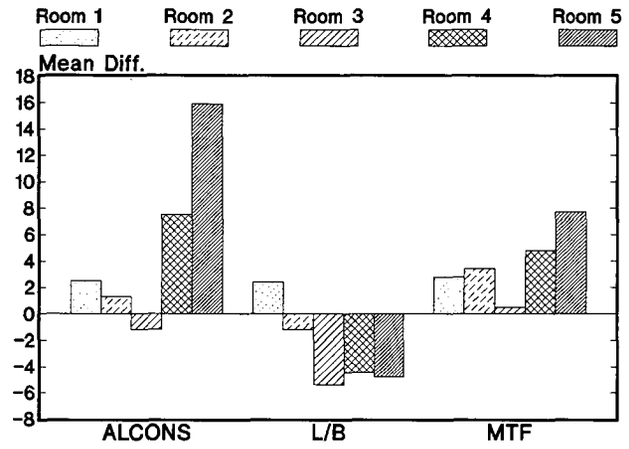


Fig. 6. Mean differences of predicted versus actual intelligibility scores.

effective signal-to-noise ratio. The positioning of the low Q source in the rooms used in this study was unfavorable for intelligibility by the same argument; the loudspeakers were in all cases positioned at the front and middle elevation of the stage, and hence were not particularly near surfaces that would be necessary to enhance early reflections.

Although the signal-to-noise and modulation transfer functions were better in predicting intelligibility, they have significant disadvantages. Foremost is that in their present form they require an impulse response as input. While methods exist for predicting an electroacoustic system's impulse response (for example ray-tracing and image-model techniques), they are difficult to implement and are computationally intensive. This presents a paradoxical situation for the sound system designer—Klein's formula is simple, but can be highly inaccurate; the signal-to-noise and modulation transfer function techniques are more accurate but are much more difficult to implement. Clearly this points the way toward future research. First, more rooms need to be characterized, particularly those with potential or real intelligibility problems. A more complete data base must be established. Second, the widespread access to computers means that predictive techniques need not be restricted to simple algebraic expressions.

7 REFERENCES

[1] 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.).

[2] J. P. A. Lochner and J. F. Burger, "The Influence of Reflections on Auditorium Acoustics," *J. Sound Vibration*, vol. 1, pp. 426–454 (1964).

[3] T. Houtgast and 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 (1985 Mar.).

[4] ANSI Std S3.2-1960 (R 1971), "Method for Measurement of Monosyllabic Word Intelligibility."

[5] K. D. Kryter, "Methods for the Calculation and Use of the Articulation Index," *J. Acoust. Soc. Am.*, vol. 34 (1962 Nov.).

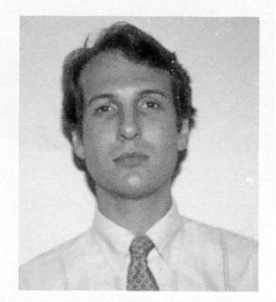
[6] V. M. A. Peutz, "Articulation Loss of Consonants as a Criterion for Speech Transmission in a Room," *J. Audio Eng. Soc.*, vol. 19, pp. 915–919 (1971 Dec.).

[7] D. Green, *An Introduction to Hearing* (Wiley, New York, 1976).

[8] H. G. Latham, "The Signal-to-Noise Ratio for Speech Intelligibility—An Auditorium Design Index," *Appl. Acoust.*, vol. 12 (1979 July).

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

THE AUTHOR



Ken Jacob is a member of the engineering staff of Bose Corporation, Framingham, MA. He received a bachelor's degree in acoustics from the University of Minnesota in 1981 and a master's degree in acoustics from the Massachusetts Institute of Technology in 1984, where his thesis topic was acoustic emission from superconducting magnets.

His interest in acoustics grew out of work as a sound designer and instructor in professional theater. He is also involved in music recording, and has produced two record albums.