Accurate Prediction of Speech Intelligibility without the Use of In-Room Measurements*

KENNETH D. JACOB, THOMAS K. BIRKLE, AND CHRISTOPHER B. ICKLER

Bose Corporation, Framingham, MA 01701, USA

The speech transmission index (STI) has been shown to be an accurate predictor of speech intelligibility in rooms, and the computationally efficient rapid STI (RASTI) method has recently become an International Electrotechnical Commission (IEC) standard. Instruments exist to measure the STI after a sound system has been installed and is operating, but until now the STI method has not been implemented and its accuracy verified in a sound system modeling program. Such an implementation has a fundamental advantage in that it does not require acoustic measurements from the room as input; this means that intelligibility can be predicted in unbuilt or inaccessible rooms solely on the basis of modeled rather than actual acoustic behavior. A new microcomputer-based implementation of the STI method is described along with the results of an experiment designed to test its accuracy. The accuracy of the new method is shown to be essentially the same as the accuracy of predictions based on in-room measurements. These results show that speech intelligibility can be predicted reliably without using acoustic measurements.

0 INTRODUCTION

One of the challenges in designing high quality sound systems for large spaces is to deliver intelligible speech reinforcement to every listener. Listeners may tolerate problems associated with frequency response, interfering noise, distortion, or sound localization, but if speech is difficult to understand, complaints are certain to result. For this reason, sound system designers need to be confident that a design will be intelligible when installed. While instruments are available which estimate the intelligibility of installed sound systems [1]–[3], no comprehensive and accurate method exists for predicting intelligibility in cases where acoustic measurements are impossible or impractical.[1]

There are three kinds of intelligibility prediction methods. The most direct require the use of screened and trained listeners as well as one of a number of different standardized word lists reproduced through the sound system under consideration [5]. A second, less direct set of methods requires that in-room measurements be made and used as input to one of several formulas for predicting intelligibility [6, p. 180]. These methods are attractive because they do not require the time, money, and expertise needed to conduct subject-based tests, but they cannot be used in situations where acoustic measurements are impossible. The third and least direct methods are those that use only a model to generate the input needed to use one of the intelligibility formulas. The three kinds of intelligibility prediction methods can be diagrammed as follows:

Predictions based on modeled behavior
⇓
Predictions of subject-based test scores from in-room measurements
⇓
Predictions of true system intelligibility from subject-based test scores
⇓
True intelligibility of system in actual use.

* Manuscript received 1990 May 7; revised 1990 November 26.

[1] van Rietschote, Houtgast, and Steeneken [4] developed a computer program for generating the speech transmission index using a ray-tracing and statistical acoustics approach. The program, while shown to be accurate for the special case of an omnidirectional source in a rectangular room, has not been developed for general-purpose use in sound system design.

One method for predicting speech intelligibility from in-room measurements has been accepted as an official standard by the International Electrotechnical Commission (IEC 268-16) [7]. The speech transmission index method has been found to be accurate in numerous independent studies (including [8]–[11]). For these reasons, the STI method was chosen for implementation in an existing computer program for designing sound systems [12].

The squared impulse response (squared sound pressure versus time) at a given receiver location is sufficient as input to the STI method. In the computer-based implementation of the STI method described in this study, the squared impulse response is divided into three parts: direct arrivals, discrete early reflections, and late arriving exponentially decaying reverberation. A computationally efficient representation of this response, called the Hybrid Energy Decay Curve (HEDC),[2] is used. The early part of the HEDC is composed of direct arrivals and reflected arrivals predicted using an image source method [13]. The late part consists of reverberation predicted using statistical reverberation theory [14, pp. 304–307].

The new STI implementation has been tested for accuracy by comparing its speech intelligibility predictions with subject-based intelligibility test scores obtained in 50 different conditions from 10 large rooms. At the same time, in-room measurements were made to obtain measured STI values. In this study, predictions of speech intelligibility made by the new computer-based STI method as well as by measured STI values are compared to the subject-based scores. In addition, predicted and measured STI values are compared directly to test the suitability of the HEDC as a substitute for the squared impulse response.

1 STI METHOD OF PREDICTING INTELLIGIBILITY

1.1 Rationale

The STI method developed by Houtgast and Steeneken [15] is based on the characteristics of actual continuous speech. In this method, continuous speech is reduced to an amplitude-modulated speech spectrum. Modulation in real speech occurs when the broad-band spectrum generated by the vocal cords is modulated into discrete speech sounds by the mouth. Houtgast and Steeneken measured the modulation spectrum of continuous speech and found that it ranged from about 0.5 to 12.5 Hz.

The basic requirement for preserving good speech intelligibility is for a speech signal to pass through an acoustic environment with its original modulation characteristics unchanged. The degree to which the sound system and room preserve the original modulation is therefore a good indication of their suitability for speech transmission. Both late-arriving reverberation and background noise, however, have the effect of reducing the original modulation.

Because the STI is calculated using modulation frequencies less than 12.5 Hz, it can be shown that early reflections arriving within a certain time do not lower the STI. In fact, in the presence of late arriving reverberation, an increase in the early reflection level increases the STI and the subject-based intelligibility scores [16]. It is implicit in the STI method that early reflections have a beneficial effect on intelligibility.

1.2 Calculating the Speech Transmission Index

In the STI method, modulation frequencies at one-third-octave intervals from 0.63 to 12.5 Hz are used. These 14 discrete modulation frequencies are used to modulate seven octave bands centered from 125 Hz to 8 kHz. The modulation index is the ratio of modulation at the output of the system to the modulation at the input. The modulation indices for the entire matrix of 14 modulation frequencies and seven octave bands are calculated in the STI method. (A simpler variant of the STI method called RASTI uses only nine modulation frequencies in two octave bands [2].)

Once the modulation index matrix has been computed, it is reduced to the single-number STI. This reduction is detailed elsewhere [17], but essentially requires converting the modulation indices to equivalent signal-to-noise ratios, which are then summed by octave band, and the resulting sums weighted and averaged. The process of calculating the STI is illustrated in Fig. 1.

1.3 Schroeder Method of Computing the Modulation Matrix

Although the modulation indices can be measured directly by comparing system-input to system-output modulation, Schroeder [18] derived the relationship between the source-to-receiver squared impulse response and the modulation index function. This relationship makes it possible to compute the modulation indices once the squared impulse response has been measured or predicted. The Schroeder formula is

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

where

$m(F)$ = modulation index as a function of modulation frequency F
$h^2(t)$ = squared impulse response.

The modulation index function is therefore proportional to the magnitude of the Fourier transform of the squared impulse response. Note that direct implementation of this equation does not properly account for the effect of background noise on the STI. The STI algorithm specifies an input spectrum that is not flat, but rather approximates the average power spectrum

[2] Hybrid Energy Decay Curve and HEDC are registered trademarks of Bose Corporation.

of the human voice. Thus the signal-to-background noise ratio as defined by the STI method would be different than the ratio found by measuring the system's impulse response. Fortunately, if the octave-band speech signal to background noise ratio is known, its effect can be added after the modulation indices have been computed using Eq. (1) [15].

1.4 Reported Accuracy of the STI Method

Steeneken and Houtgast [17] tested the accuracy of the STI method using a variety of conditions. These included conditions where bandpass limiting, background noise, peak clipping, automatic gain control, and reverberation were responsible for degrading intelligibility. They scored speech intelligibility using trained listeners and lists of Dutch monosyllabic nonsense words. A third-order regression function was fit to the measured STI versus subject-based speech intelligibility data, and a standard deviation about this regression function of $\sigma = 5.6\%$ was found. In practical terms this means that intelligibility can be predicted to within $\pm 5.6\%$ most of the time. When their analysis was confined to distortions in the time domain (background noise, reverberation, and automatic gain control), the standard deviation of the data about the regression curve was similar at $\sigma = 5.8\%$.

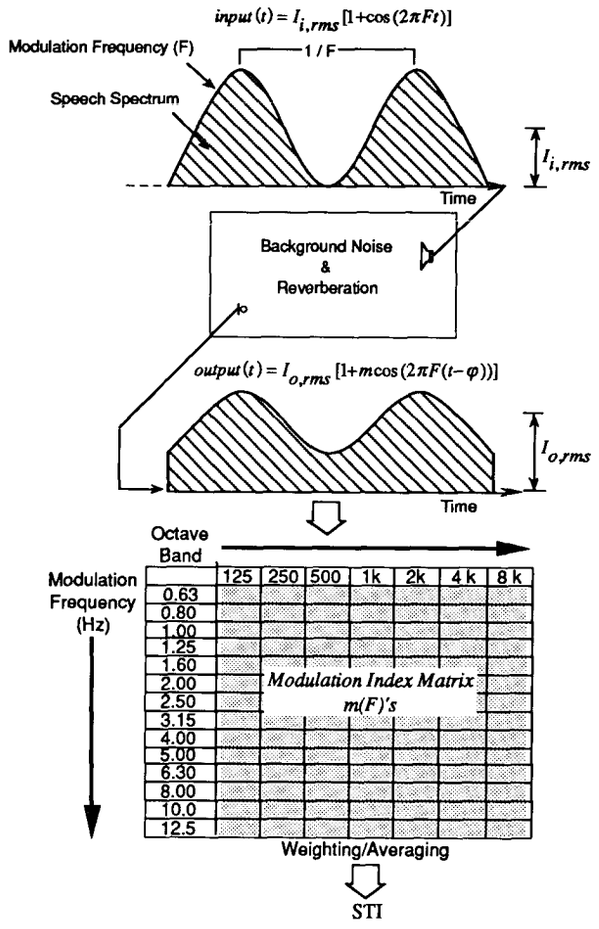


Fig. 1. In the STI method, an artificial signal modeled after actual speech is used. This signal consists of amplitude-modulated speech spectrum as shown. Background noise and reverberation both have the effect of reducing the modulation present in speech, and the modulation index m is a measure of the modulation loss as the signal passes through the sound system and room. The modulation index matrix is a matrix of seven octave bands and 14 modulation frequencies. From the matrix, the single-number STI is computed.

2 COMPUTER IMPLEMENTATION OF STI METHOD

2.1 Room and Sound System Modeling

The STI method has been integrated into a computer program for predicting the performance of sound systems in rooms [12]. Rooms are modeled in the program by drawing a series of N-sided planes (where $N \leq 10$), each of which is assigned a surface material defined by octave-band Sabine absorption coefficients. Sound sources are represented by their full-space octave-band directional responses from 125 Hz to 4 kHz, and by their sensitivity and input power. Sound systems are modeled by specifying cluster locations, source types, source orientation, source powers, and any electronic time delays. An example of one of the rooms modeled in this study is shown in Fig. 2.

The program uses three algorithms to produce output relevant to the sound system designer. Direct field contributions are predicted by adding source directional attenuation (if any) to the source-to-listener inverse square loss. Discrete early reflections are predicted using an image source method [13], and late-arriving reverberation is predicted using the classic formulas of statistical room acoustics [14].

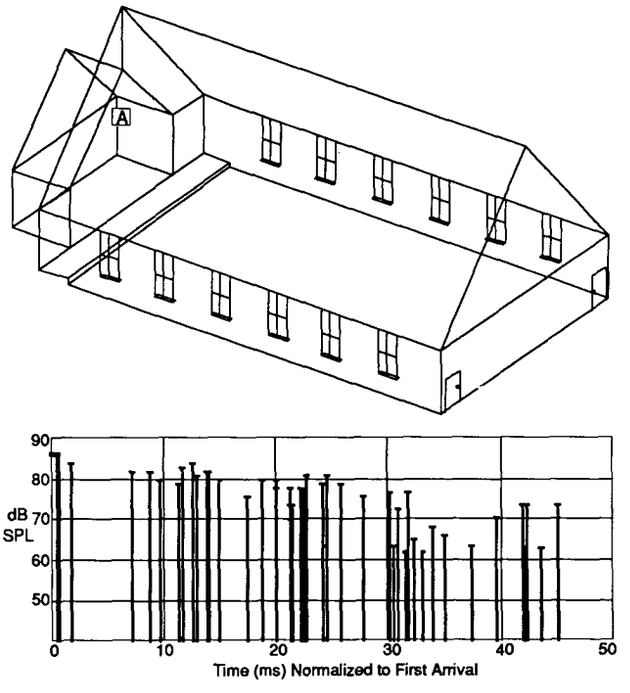


Fig. 2. Medium-sized church as modeled in computer program [12]. Rooms are constructed as a series of adjacent planes, each of which is assigned a surface material. Sound sources are represented by their octave-band full-space polar responses. The early time response (consisting of direct field and early reflections) is shown for a low-Q loudspeaker located at position A and a receiver in the rear of the room.

2.2 Hybrid Energy Decay Curve (HEDC)—Rationale

As discussed in Sec. 1.3, the modulation index matrix needed to find the STI can be computed using Eq. (1), which requires the squared impulse response as input. Prediction of the squared impulse response requires the prediction of a large number of reflections. On average, it can be shown [6, p. 87] that the number of reflections arriving within a certain time after excitation of a source is

$$N = \frac{4\pi c^3 t^3}{3V} \tag{2}$$

where

N = number of reflections arriving within time t after excitation of source
c = speed of sound
V = volume of the room.

As an example, in an auditorium of $V = 6000$ m^3, 28 reflections arrive in the first one-tenth of a second, 3522 arrive in the first one-half second, and 28,172 arrive in the first second. Therefore most of the calculations required to predict the source-to-receiver squared impulse response are associated with the large number of reflections that occur in typical rooms.

Eq. (2) can be differentiated with respect to time to obtain the average number of reflections per unit time, or the reflection density function,

$$\frac{\partial N}{\partial t} = \rho(t) = \frac{4\pi c^3 t^2}{V} \tag{3}$$

where $\rho(t)$ is the number of reflections per unit time.

Eq. (3) shows that the reflection density increases with time squared. In other words, the room rapidly becomes filled with a great number of wave fronts whose behavior typically becomes more and more random with time. Under these conditions, it is possible to describe the behavior using statistics—to average the wave fronts instead of attempting to follow them independently. Accounting for the average behavior of reverberation is computationally much simpler than accounting for each individual reflection.

In generating the Hybrid Energy Decay Curve, an image source method is used to predict discrete early arrivals which have undergone three or fewer reflections; these are the arrivals that are uniquely characteristic of the sound system (loudspeaker types, loudspeaker aiming angles, power levels, and delay) and the room (geometry and specific distribution of absorption). A statistical model of reverberation is then used to attach a late exponentially decaying "tail" to the early part. The result is a representation of sound transmission which exploits the sophistication of the model when it is still computationally efficient to do so, and then switches to a much simpler description to account for the remaining late statistical reverberation. A graphic representation of the HEDC is shown in Fig. 3.

2.3 Converting the HEDC to the Modulation Index Matrix

The modulation index matrix needed to find the STI is computed from the HEDC using a modification of Eq. (1). The two parts of a given octave-band HEDC are each transformed separately as follows:

$$m(F) = \left| \frac{\sum_{i=1}^{I} \int_0^\infty r_i \delta(t - t_i) e^{-j2\pi Ft}\, dt + \int_{t_d}^\infty A e^{-b(t-t_d)} e^{-j2\pi Ft}\, dt}{\int_0^\infty \mathrm{HEDC}(t)\, dt} \right| \tag{4}$$

where

$m(F)$ = modulation index function (also known as modulation transfer function)
I = total number of discrete early arrivals
r_i = squared pressure level of ith discrete early arrival
t_i = time of arrival of ith discrete early arrival
A = initial level of late exponentially decaying reverberation
b = decay rate (time constant of decay)
t_d = time at which late reverberation begins.

3 EXPERIMENT TO TEST ACCURACY OF NEW METHOD

3.1 Rooms, Sound Sources, and Listener Positions

Ten rooms (all in the Boston metropolitan area), three sound sources, and two listener positions per room

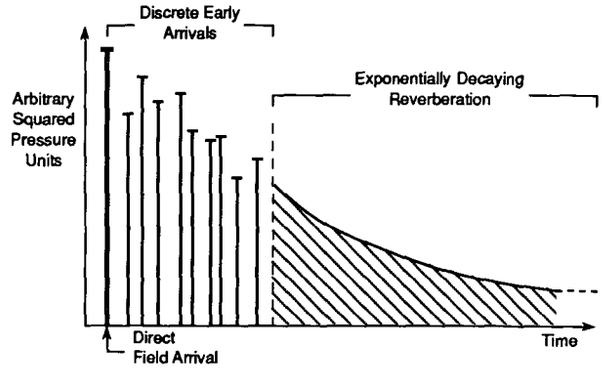


Fig. 3. Graphic representation of the Hybrid Energy Decay Curve. The HEDC consists of an early part, composed of direct field and early reflections predicted using an image source method, and a late part, consisting of exponentially decaying reverberation and predicted using classical reverberation theory.

comprised a database of 50 different conditions for speech intelligibility. The details of these conditions are described extensively elsewhere [11]. The rooms ranged in size, architectural complexity, and reverberation characteristics. The sources were chosen for their wide range of directional characteristics, and listener positions were chosen to represent positions both near to and far from the sources. These conditions are summarized in Tables 1–3.

3.2 Subject-Based Testing

Intelligibility tests using subjects were carried out in each of the 50 conditions; the exact details are described elsewhere [11]. The tests were administered according to ANSI S3.2-1971 [19]. Subject-based intelligibility scores are denoted by %PB-ansi in this study. (PB refers to the fact that the words are phonetically balanced to match those of normal language usage.) The total number of words presented in each condition ranged from 2000 to 2800. Mean scores for the 50 conditions are given in the Appendix.

3.3 In-Room Measurements of the STI

For each room, source, and listener position combination, system impulse responses were recorded. The STI for each of these measured impulses was computed by applying Eq. (1) and reducing the resulting modulation index matrix to the STI according to the procedure defined by Houtgast and Steeneken [17]. These values are denoted by STI-measured and are tabulated in the Appendix.

Table 1. Rooms.

Name	T_{60}*	Function
Berklee Performance Center	0.9	Music
Coolidge Corner Movie House	1.0	Cinema
Huntington Theater	1.1	Drama
Saint Bridget's Church	2.0	Religious
Nevins Hall	3.5	Multifunction
Jordan Hall	2.2	Music
Mechanics Hall	2.2	Music
South End Cathedral	3.3	Religious
Cyclorama	3.5	Multifunction
MIT Indoor Track	4.6	Athletics

* Reverberation times are averages of the measured times in the 1-, 2-, and 4-kHz octave bands.

3.4 Room Modeling and STI Prediction

Computerized room models were created for each of the 10 rooms. Architectural details such as columns and stairs were simplified in order to reduce the number of planes; the result was that less than 100 planes were used in each room model. Materials were chosen from a standard list [14, p. 300], and source and receiver locations were entered into the computer to match their actual locations. For each room, source, and listener position combination, an HEDC was calculated in each octave band from 125 Hz to 4 kHz. Each HEDC was transformed using Eq. (4), and the results were used to complete the modulation index matrix. (The 8-kHz data in the matrix were copied from the 4-kHz data, but were later weighted differently according to the method used to reduce the modulation matrix specified by the STI method.) Finally, the modulation index matrix was reduced to the STI; these values are denoted by STI-predicted and are tabulated in the Appendix.

4 RESULTS

Data were correlated three ways. First, the relationship between the in-room measurements of the STI and the subject-based speech intelligibility scores was found. These data reveal the basic accuracy of the STI method. Second, the relationship between predicted STI values (from the computer model) and the same subject-based scores was found. This relationship shows the accuracy of the new computer-based implementation of the STI. Third, the relationship between the measured and predicted STI values was found. This direct comparison, while not in itself correlated with actual subject-based intelligibility, does show the ability of the HEDC to represent the source-to-receiver squared impulse response.

Table 2. Loudspeakers.

Name	Type	Directivity*
Soundsphere 2212-1	Omniradiator	1
Bose 802-II	Eight-driver array	8
Electro Voice HR6040A	Constant-directivity horn†	18

* Loudspeaker directivities are averages of the measured on-axis directivities in the 1-, 2-, and 4-kHz octave bands.

† The horn loudspeaker was used in conjunction with an Electro-Voice TL806AX bass-to-midrange loudspeaker, thereby completing a full-spectrum system.

Table 3. Listener Positions.

Name	Relationship to source	Position in room
Near position	On axis ±7.5°	One third of room length
Far position	On axis ±7.5°	Rear of room

4.1 Relationship between In-Room Measurements of the STI and Subject-Based Scores

A third-order polynomial regression function was computed for the subject-based speech intelligibility versus measured STI data. The regression function and a scatter plot are shown in Fig. 4. The standard deviation of the data about the regression curve is 5.2%, which is similar to the 5.8% value reported in Steeneken and Houtgast [17]. The regression function is

$$\%\text{PB-ansi} = 788.26\text{STI}^3 - 1643.9\text{STI}^2 + 1179.3\text{STI} - 196.3 \ . \quad (5)$$

4.2 Relationship between Predictions of STI and Subject-Based Scores

A scatter plot of subject-based intelligibility data from the ANSI test and predicted STI values is shown in Fig. 5 along with the regression function of Eq. (5). The standard deviation of these data about the regression function is 5.4%. Thus the overall error in predicting speech intelligibility using predicted STI values is essentially equivalent to the error using measured STI values. This means that there is no significant penalty for moving from the domain of in-room acoustic measurements to that of a pure computer model in terms of predicting speech intelligibility using the STI method.

4.3 Relationship between Measured and Predicted STI values

While the preceding results show that speech intelligibility can be estimated with the same accuracy from predicted STI values as from measured STI values, it is also of interest to study the direct relationship between the measured and predicted STI values. A scatter plot showing this relationship is given in Fig. 6. The correlation coefficient for the data is $r = 0.81$ and the standard deviation of the data is $\sigma = 0.06$, which is considered good to very good.[3] This result shows that the HEDC is a reasonable substitute for the actual squared impulse response when used to predict speech intelligibility using the STI method.

5 DISCUSSION

5.1 Overall Accuracy of the STI Method

Results show that the predictions of speech intelligibility using the new computer-based implementation of the STI method are essentially as good as those based on in-room measurements of the STI. This is an important step in providing to the sound system designer a tool for predicting speech intelligibility in unbuilt or inaccessible rooms. However, it is important to note the significance of a standard deviation of 5–6%. The Steeneken and Houtgast data [17] and the measured and predicted STI data from this study all show standard deviations in this range. A standard deviation of 5–

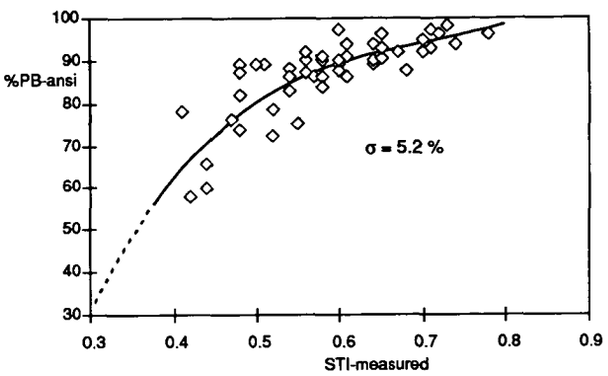


Fig. 4. Scatter plot of speech intelligibility scores from ANSI test versus measured STI values (from the 50-condition database) along with third-order regression function best fitting the data. The error of 5.2% found for these data is about the same as that found by Steeneken and Houtgast [17].

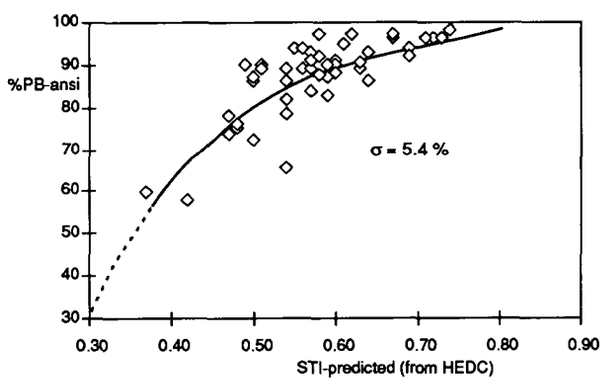


Fig. 5. Scatter plot of speech intelligibility scores from ANSI test versus predicted STI values (from computer-generated HEDCs) along with third-order regression function of Fig. 4. Notice that the fit of the data to the regression curve is essentially the same as in Fig. 4.

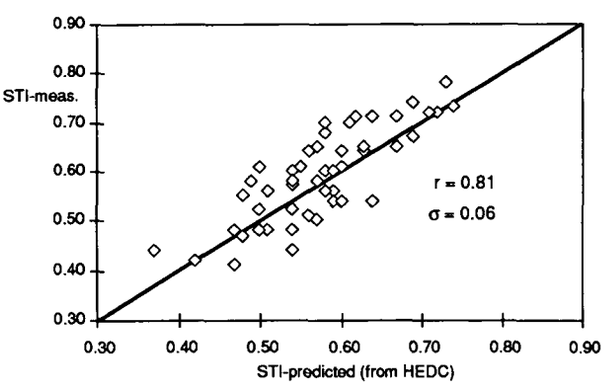


Fig. 6. Scatter plot of measured versus predicted STI values. The straight line represents perfect correlation. The correlation coefficient for these data of $r = 0.81$ is considered good to very good.

[3] Note that this standard deviation is given in STI units and should not be compared numerically to previous standard deviations, which have all been in percent word intelligibility units.

6% means that the subject-based intelligibility scores will be within 5–6% of the predicted intelligibility scores most of the time. This inherent error must be included in interpreting predictions of speech intelligibility based on the STI, even when it is measured.

The 5–6% error inherent in the STI method, while no greater than some published methods and much less than others [11], may be reduced in future studies. For example, the STI method currently does not weight modulation frequencies. It may be that some modulation frequencies are more important than others, such as those responsible primarily for producing the consonant sounds.

The STI method is a single-channel method. Yet there is clear evidence [20] that binaural hearing has a dramatic effect on speech intelligibility. For example, listeners using both ears get much higher intelligibility scores than those using one ear or those using headphones supplied with an identical signal to both ears. It is possible that the inherent error in the STI method could be reduced if these binaural effects were considered.

Lastly, there is some evidence [21], [22] that one-third-octave resolution coupled with a more drastic spectral weighting function could improve the accuracy of the STI method. However, both loudspeaker and room material data currently only exist in one-octave-band resolution. One-third-octave resolution would increase greatly the computational and memory requirements needed to predict the STI from a computer model.

5.2 Limitations of the Computer-Based STI Method

The microcomputer-based implementation of the STI method described in this study is based on an image-source method for predicting early reflections and statistical room acoustics theory to predict late exponentially decaying reverberation. Both models have known limitations. In the image model it is assumed that a room boundary can always be approximated by a flat plane, which reflects sound in the same way as a mirror reflects light. Real surfaces, however, can both reflect and scatter sound waves, thereby changing the means by which sound energy is transmitted to the listener. Modeled surfaces are simply assigned an octave-band absorption coefficient which absorbs sound by the same amount, regardless of the angle at which the sound wave strikes the surface. Real surfaces exhibit absorption properties which are a function of incident wave angle. Finally, modeled surfaces are assumed to be large compared to a wavelength of sound. At low frequencies this is not necessarily the case. These simplifications result in predictions of discrete early reflections that can deviate substantially from reality.

While the prediction of each individual reflection is subject to error, there appears to be no systematic error applying to all reflections. In the HEDC-based STI method described here, all of the predicted reflections are used in the computation, and their large number tends to reduce the error that would result if fewer reflections were predicted. In this study the rooms used were typical (see Table 1). However, it is possible that rooms with many surfaces which are known to violate the assumptions of the image-model method (a room covered with purposely diffusing panels, for example) would result in higher errors.

In the case of the model used to predict late arriving exponentially decaying reverberation, the assumptions are essentially that sound absorption is distributed evenly throughout the room and that the sound field is diffuse. Real rooms, of course, do not always obey these assumptions. Many rooms do not have evenly spread absorption. (An auditorium with hard surfaces except for the seating areas is typical.) In these cases, either the strength or the decay rate of the late arriving reverberation can be predicted incorrectly using classical formulas. The assumption of sound field diffuseness is probably easier to meet in real rooms, since the rapid buildup of wave fronts with time ensures rapid convergence to diffuse conditions [see Eq. (3)]. In the new method described here, errors in the prediction of late reverberation would appear in the late part of the HEDC and therefore in the STI. The correlation between both the predicted STI values and actual subject-based intelligibility, and the predicted and measured STI values suggest, however, that this error is not limiting in terms of accurate prediction of speech intelligibility using the STI method.

The 50 conditions used in this study each represent instances where reverberation is responsible for degrading speech intelligibility. Background noise was minimized as a factor by guaranteeing in each condition that the speech-signal-to-background-noise ratio exceeded 15 dB [11]. While this emphasis on the effect of reverberation was intentional in order to test the suitability of the Hybrid Energy Decay Curve, the effect of background noise was not tested explicitly. This effect has been studied extensively by Houtgast and Steeneken [15], [17], [23] and others (including [9], [10], [24]).

Last, the database in this study, while large, does not include examples of some typical sound system types. Purely distributed systems and systems using loudspeakers with electronic delay have not been tested. However, no additional limitations should exist in generating the HEDC for these system types than already exist for the system type included in this study.

5.3 Relationship between the STI and Subject-Based Tests

The relationship established between in-room measurements of the STI and subject-based intelligibility scores as measured using the ANSI method [Eq. (5) and Fig. (4)] is unique to this study. Other studies used different subject-based tests, methods of training subjects, or test conditions, and have therefore arrived at different regression functions relating the STI to specific subject-based speech intelligibility tests. The details of some of these relationships and a discussion of the differences are given in the Appendix.

5.4 Error Analysis

There is an error in predicting subject-based scores from in-room measurements of the STI, as shown in Fig. 4. There is about the same error in predicting subject-based scores from modeled (predicted) STI values, as shown in Fig. 5. It must be stressed that STI values generated by the new computer-based method are used to predict subject-based scores *directly* using Eq. (5), and are used only secondarily as a comparison to measured STI values. The fact that the predicted STI values are not exactly the same as the measured STI values (Fig. 6) does not imply that predicted STI values are poorer predictors of subject-based scores. No "double error" occurs in predicting intelligibility from the modeled STI values since they are not first converted to measured STI values.

6 CONCLUSION

The STI method of estimating speech intelligibility has been implemented in a microcomputer-based program for predicting sound system performance. The new STI implementation requires no in-room measurements and is thus suitable for unbuilt or inaccessible rooms. The new technique relies on a computationally efficient method of representing the transmission of sound from a sound source to a listener, called the Hybrid Energy Decay Curve (HEDC). This hybrid curve is generated using an image source method to predict discrete early arrivals and statistical acoustics theory to predict late arriving exponentially decaying reverberation.

Results of an experiment designed to test the accuracy of the new implementation show that no loss of accuracy occurs in predicting speech intelligibility when compared to predictions based on in-room measurements of the STI. The accuracy of speech intelligibility predictions was shown to be $\pm 5.4\%$. These results mean that sound system designers can predict with known accuracy the speech intelligibility of unbuilt or in-process designs. The results also show that the correlation between predicted STI values and measured STI values is good to very good ($r = 0.81$). This demonstrates that the HEDC is a good substitute for the actual source-to-receiver squared impulse response in this application.

The overall accuracy of the STI method was shown to be important in terms of predictions of speech intelligibility. While predictions based on either measured or modeled STI values are as good, and in some cases much better than other methods, sound system designers should be mindful that predictions can only be described as good, not excellent.

Last, it was the intent of the authors to describe this study in sufficient detail for it to be reproduced by others. However, repetition requires the use of the computer program in which the STI method has been implemented. Investigators wishing to reproduce this experiment or some variant of it should contact the authors in order to receive permission to use the computer program.

7 REFERENCES

[1] H. Steeneken and E. Agterhuis, "Description of STIDAS II-D: General System and Program Description," Rep. IZF 1982-29 Inst. for Perception TNO, 1982.

[2] "RASTI," *Brüel and Kjær Tech. Rev.*, no. 3, 1985.

[3] D. Keele, Jr., "Evaluation of Room Speech Transmission Index and Modulation Transfer Function by the Use of Time Delay Spectrometry," in *Proc. AES 6th Int. Conf.* (Nashville, 1988).

[4] H. van Rietschote, T. Houtgast, and H. Steeneken, "Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function IV: A Ray-Tracing Computer Model," *Acustica*, vol. 49, no. 3 (1981).

[5] W. Voiers, "Uses, Limitations, and Interrelations of Present-Day Intelligibility Tests," presented at the National Electronics Conference (Chicago, 1981 Oct).

[6] H. Kuttruff, *Room Acoustics*, 2nd ed. (Appl. Science Publ., London, 1979).

[7] IEC 268-16, "The Objective Rating of Speech Intelligibility in Auditoria by the RASTI Method," 1st ed., International Electrotechnical Commission, Geneva, Switzerland (1988).

[8] H. Nomura et al., "Speech Intelligibility and Modulation Transfer Function in Non-Exponential Decay Fields," *Acustica*, vol. 69 (1989).

[9] B. Anderson and J. Kalb, "English Verification of the STI Method for Estimating Speech Intelligibility of a Communications Channel," *J. Acoust. Soc. Am.*, vol. 81, no. 6 (1987).

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

[11] K. D. Jacob, "Correlation of Speech Intelligibility Tests in Reverberant Rooms with Three Predictive Algorithms," *J. Audio Eng. Soc.*, vol. 37, pp. 1020–1030 (1989 Dec.).

[12] Modeler Design Program v3.1.

[13] J. Borish, "Extension of the Image Model to Arbitrary Polyhedra," *J. Acoust. Soc. Am.*, vol. 75, no. 6 (1984).

[14] L. Beranek, *Acoustics* (Am. Inst. Phys. for Acoust. Soc. Am., New York, 1986).

[15] T. Houtgast and H. 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).

[16] K. Jacob, "The Role of Early and Late Reflections in the Prediction of Speech Intelligibility," presented at the 4th *Reg. Conf. of the Audio Eng. Soc.* (Tokyo, Japan, 1989 June).

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

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

[19] ANSI S3.2-1971, "Standard for Measuring Monosyllabic Speech Intelligibility," Am. Nat. Standards Inst., New York, 1971.

[20] E. Carterette and M. Friedman, Eds., *Handbook of Perception*, vol. IV: *Hearing* (Academic Press, New York, 1978), pp. 444–446.

[21] L. Humes et al., "Application of the Articulation Index and the Speech Transmission Index to the Recognition of Speech by Normal-Hearing and Hearing-Impaired Listeners," *J. Speech Hear. Res.*, vol. 29 (1986 Dec.).

[22] L. Humes et al., "Further Validation of the Speech Transmission Index," *J. Speech Hear. Res.*, vol. 30 (1987 Sept.).

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

[24] K. Kryter and E. Whitman, "Some Comparisons between Rhyme and PB-Word Intelligibility Tests," *J. Acoust. Soc. Am.*, vol. 37, p. 1146 (1965).

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

[26] T. Houtgast and H. Steeneken, "A Multi-Language Evaluation of the RASTI Method for Estimating Speech Intelligibility in Auditoria," *Acustica*, vol. 54 (1984).

[27] A. Mochimaru, "An Evaluation of the Accuracy of MTF-STI Measurements by Comparison to Japanese Three-Syllable Listening Tests," presented at the 89th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 38, p. 868 (1990 Nov.), preprint 2941.

APPENDIX
Relationship between Measured STI and Various Subject-Based Tests

In this study, phonetically balanced (PB) monosyllabic English words were used as specified by ANSI (S3.2-1971) [19]. The relationship between the STI and the ANSI test, shown as the regression function of Eq. (5) and Fig. 4, is unique to this study. Other studies have used different subject-based tests, and the relationships they have established between the STI and these other tests are therefore different. Because of these differences, the subject-based test must be specified in converting the STI to speech intelligibility; the method of testing, the type of words, the presence or lack of a carrier sentence to embed the words, and the language all affect the relationship of the STI to test scores. Some of the other studies relating the STI to subject-based tests are compared here.

Houtgast, Steeneken, and Plomp used phonetically balanced Dutch consonant–vowel–consonant (CVC) nonsense words [23]. Bradley [10] measured intelligibility using the Fairbanks rhyme test, which is a multiple-choice test in which the words in each set rhyme. Anderson and Kalb [9] related the STI to two types of speech intelligibility tests. For one, they used the same words as this study, but "thoroughly familiarized" their listeners with them. In the second they estimated the relationship of the STI to the modified rhyme test (a variant of the Fairbanks rhyme test) by using data from another study [24]. The relationships between the STI and these various subject-based tests are shown in Fig. 7.

Inspection of the regression functions of Fig. 7 shows that the relationship between the STI and speech intelligibility established in this study is similar to that established by Houtgast and Steeneken for Dutch CVC nonsense words. The similarity may be interpreted as saying that using English words and less subject training is approximately the same as using nonsense words and more training. However, there appears to be a penalty in using less training—the need to use a much larger number of words to achieve similar accuracy. In this study, each data point represents the transmission of between 2000 and 2800 words, whereas Houtgast and Steeneken used only 400 [25] for each of their conditions. All else being equal, this represents a factor of 5–7 difference in testing time.

Fig. 7 also shows that subjects score much higher on the two rhyme tests than in the other subject-based tests. This is expected because rhyme tests are multiple-choice (closed-set) tests. In addition, in the rhyme tests, words differ only in one consonant. CVC words contain by definition two consonants, and PB words average almost two per word, making the chances for error about twice as high as in the rhyme tests.

The Anderson and Kalb function[4] shows that their subjects scored higher than those of this study or those of the Houtgast and Steeneken study. One possible explanation is that Anderson and Kalb "thoroughly familiarized" their subjects with the words before be-

[4] The Anderson and Kalb regression function for PB words, shown in Fig. 7, is derived from the data in [9, Fig. 1] since the equation quoted by the authors for this regression function appears to be in typographical error.

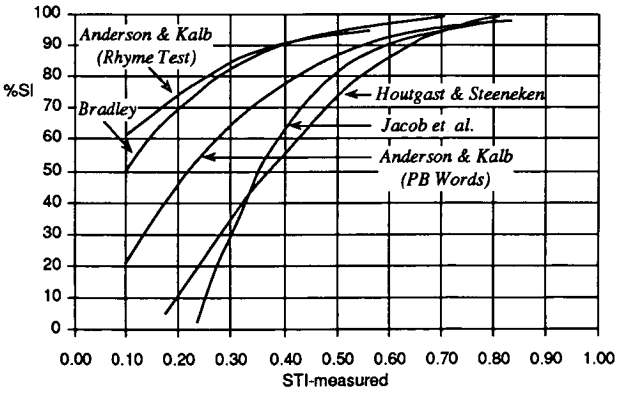


Fig. 7. Relationships between various subject-based intelligibility tests used in this and other studies and STI. The differences are primarily the result of different test types. The top two curves are from rhyme tests where higher scores are obtained because the tests are multiple choice and the words differ only by one consonant. The two middle curves are from tests using monosyllabic English words. The lowest curve (from Houtgast and Steeneken) is from a test using monosyllabic nonsense words.

ginning testing. To some degree their subjects may have been choosing from a closed set. Under these conditions, an increase in the intelligibility scores would be expected. Another possible explanation is that Anderson and Kalb used single-channel artificial reverberation, whereas Houtgast and Steeneken and this study used reverberation from real rooms. In addition Anderson and Kalb only used one reverberant decay rate and used (without explanation) an initial delay in its onset of 95 ms. It is possible therefore, that they inadvertently created anomalous conditions for which the STI measure was not specifically developed.

These results show that the relationship between the STI and speech intelligibility is strongly dependent on the type of subject-based speech test used. The test used by the originators of the STI method was shown to be similar to the ANSI method used in this study. Closed-set rhyme tests or special training of subjects leads to significantly different relationships to the STI measure. These results also point out the need to establish the relationship between the STI and other languages, although some preliminary work has been conducted in this area [26], [27].

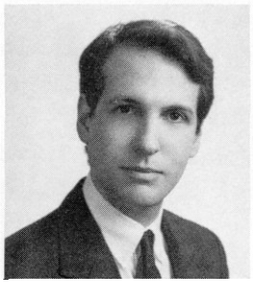
MEASURED AND PREDICTED STI VALUES, AND SUBJECT-BASED SPEECH INTELLIGIBILITY SCORES

Exact experimental details of subject-based testing can be found in [11]. In Table 4 STI-meas. refers to in-room measurements of the STI; STI-pred. refers to predicted STI values from the computer program, and %PB-ansi refers to the mean score on intelligibility word lists.

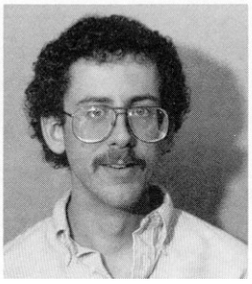
Table 4. Measured and predicted STI values and subject-based speech intelligibility scores.

Room	Source	Position	STI-meas.	STI-pred.	%PB-ansi
Berklee	Sphere	Near	0.65	0.67	96
		Far	0.71	0.64	93
	Array	Near	0.72	0.72	96
		Far	0.72	0.71	96
	Horn	Near	0.73	0.74	98
		Far	0.78	0.73	96
Coolidge	Sphere	Near	0.60	0.58	97
		Far	0.56	0.51	90
	Array	Near	0.71	0.62	97
		Far	0.64	0.56	94
	Horn	Near	0.71	0.67	97
		Far	0.61	0.60	91
Huntington	Sphere	Near	0.61	0.55	94
		Far	0.57	0.54	86
	Array	Near	0.70	0.61	95
		Far	0.64	0.63	89
	Horn	Near	0.74	0.69	94
		Far	0.67	0.69	92
Bridget's	Sphere	Near	0.56	0.58	92
		Far	0.48	0.54	82
	Array	Near	0.70	0.58	92
		Far	0.54	0.60	88
	Horn	Near	0.65	0.57	93
		Far	0.54	0.64	86
Nevins	Sphere	Near	0.41	0.47	78
		Far	0.48	0.51	89
	Array	Near	0.48	0.50	87
		Far	0.51	0.56	89
	Horn	Near	0.50	0.57	89
		Far	0.60	0.59	90
Jordan	Array	Near	0.60	0.54	89
		Far	0.52	0.54	78
	Horn	Near	0.64	0.60	90
		Far	0.56	0.59	87
Mechanics	Array	Near	0.58	0.54	86
		Far	0.54	0.59	83
	Horn	Near	0.60	0.58	87
		Far	0.65	0.63	91
Cathedral	Array	Near	0.58	0.49	90
		Far	0.47	0.48	76
	Horn	Near	0.58	0.57	91
		Far	0.44	0.54	66
Cyclorama	Array	Near	0.61	0.50	86
		Far	0.48	0.47	73
	Horn	Near	0.68	0.58	87
		Far	0.52	0.50	72
MIT Track	Array	Near	0.55	0.48	75
		Far	0.44	0.37	60
	Horn	Near	0.58	0.57	84
		Far	0.42	0.42	58

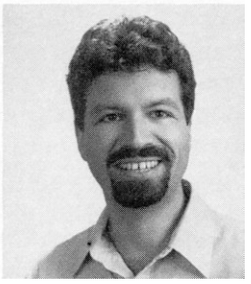
THE AUTHORS



K. D. Jacob



T. K. Birkle



C. B. Ickler

Ken Jacob is a staff engineer at Bose Corporation, Framingham, MA. 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. Mr. Jacob joined Bose in 1984 and is currently manager of acoustic research.

•

Tom Birkle is a staff engineer at Bose Corporation, Framingham, MA. He received a B.S.E.E. degree from the University of Colorado at Boulder in 1982. From 1983 to 1987 he worked as a sound system designer with the acoustical consulting firm of David L. Adams Associates, Inc. in Denver, CO. He joined Bose in 1987, where he has authored the Sound System Software family of computer programs, including the Modeler[1] Design Program.

•

Chris Ickler is a staff engineer at Bose Corporation, Framingham, MA. He received his bachelor's degree in physics from the Massachusetts Institute of Technology in 1979. In 1980 he joined Bose Corporation, where he designed a number of home high-fidelity loudspeakers. In 1985 he joined the acoustic research group where his work has been in psychoacoustics, computerized measurement and processing techniques, and mathematical modeling of acoustic phenomena.

[1] Trademark of Bose Corporation.