Accurate Prediction of Speech Intelligibility without the use of In-Room Measurements

Preprint# 2942
Session-Paper# B-3

Kenneth D. Jacob, Thomas K. Birkle, and Christopher B. Ickler
Bose Corporation
Framingham, MA USA

Presented at
the 89th Convention
1990 September 21–25
Los Angeles

This preprint has been reproduced from the author's advance manuscript, without editing, corrections or consideration by the Review Board. The AES takes no responsibility for the contents.

Additional preprints may be obtained by sending request and remittance to the Audio Engineering Society, 60 East 42nd Street, New York, New York 10165 USA.

AN AUDIO ENGINEERING SOCIETY PREPRINT

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

The Speech Transmission Index (STI) has been shown to be an accurate predictor of speech intelligibility in auditoria, and the computationally more efficient RASTI method has recently become an official International Electrotechnical Commission (IEC) standard. While instruments have been developed by others to measure the STI after room construction is complete and the sound system is operating, until now the STI method has not been implemented and tested 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 behavior. In this study, 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 accurately predicted 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, ambient noise, and localization, but if the 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, 2, 3], no comprehensive and accurate system has been developed 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, and one of a number of different standardized word lists reproduced through the sound system under consideration [5]. Another, 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]. These methods are attractive because they

[1] Rietschote, Houtgast, and Steeneken [4] developed a computer program for generating the STI 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 .

do not require the time, money, and expertise needed to conduct subject-based tests, but are by definition not applicable to cases where acoustic measurements are not possible. Least direct are methods which use only a model to generate the input needed to use one of the intelligibility formulas. The three kinds of intelligibility prediction methods are illustrated below:

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

One algorithm for predicting subject-based test scores from in-room measurements has been accepted as an official standard by the International Electrotechnical Commission (IEC #268-16). The Speech Transmission Index method has been found to be accurate in numerous independent studies [including 7, 8, 9,10]. For these reasons, the STI method was chosen for implementation in an existing computer program for designing sound systems [11].

The STI method requires the squared impulse response, or sound energy vs. time response at a given listener location as input. A computer-based implementation of the method must therefore be capable of generating this function. In the implementation described in this study, the energy vs. time response is divided into three parts: direct arrivals, discrete early arrivals reflected from room boundaries, and late diffuse reverberation. A computationally efficient representation of this response, called the Hybrid Energy Decay Curve (HEDC), is used. The HEDC [12] is composed of an early part, consisting of direct arrivals and reflected arrivals predicted using an image source method [13], and a later part consisting of late reverberation predicted using statistical reverberation theory [14]. The result is an accurate representation of the essential information needed as input to the STI algorithm.

The new STI implementation has been tested for accuracy by comparing its speech intelligibility predictions with subject-based intelligibility test scores obtained in fifty different conditions from ten auditoria. At the same time, in-room measurements were made to obtain measured STI values.

Predictions of speech intelligibility made by the new computer-based STI method as well as by measured STI values were compared to the subject-based scores. In addition, predicted and measured STI values were compared to test the suitability of the HEDC as a substitute for the actual source-to-receiver energy vs. time response.

1. STI METHOD OF PREDICTING INTELLIGIBILITY

1.1 Rationale

The STI method developed by Houtgast and Steeneken [15] models the behavior of actual speech. In this model, continuous speech is reduced to an amplitude-modulated speech spectrum. The modulation occurs when the broadband spectrum created by the vocal cords is transformed 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 to 12.5 Hz.

The general requirement for preservation of intelligibility from talker to listener is for the speech signal to pass through the acoustic environment with the original modulation characteristics unchanged. Late-arriving reflections and background noise both have the effect of reducing the amount of modulation present in the original speech signal. Thus the degree to which the sound system and room combination preserves the original modulation is a good indication of its suitability for speech transmission.

The Modulation Transfer Function (MTF) quantifies the modulation reduction of a speech transmission system. As an example, Fig. 1 shows the effects late reverberation and background noise have on one modulation frequency; as they increase, the modulation, as measured by the modulation index, is reduced. Perfect transmission results in a modulation index of 1.0. Complete loss of the original modulation results in a modulation index of 0.

Because the Speech Transmission Index 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 reflections, an increase in early reflection level increases the STI and the subject-based intelligibility scores [16]. It is central to the STI method that early reflections have a beneficial effect on intelligibility.

In addition to providing the information necessary to compute the Speech Transmission Index, the modulation transfer function contains useful diagnostic information. The acoustic conditions responsible for reducing the modulation indices can often be determined by simple inspection of a speech system's modulation transfer function. Examples illustrating these diagnostic functions are shown in Fig. 2.

1.2 Computing the Speech Transmission Index

In applying the modulation transfer function to the problem of predicting speech intelligibility, modulation frequencies at one-third octave intervals from 0.63 to 12.5 Hz are used. These fourteen discrete modulation frequencies are used to modulate seven octave bands centered from 125 Hz to 8 kHz. The STI method requires the calculation of the modulation indices for the entire matrix of fourteen modulation frequencies in seven octave bands. A simpler method called RASTI (RApid STI) uses only nine modulation frequencies in two octave bands [2].

Once the modulation transfer functions have been computed for each of the seven octave bands, they are reduced to the single number Speech Transmission Index. This reduction is detailed elsewhere [15] 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.

1.3 Schroeder Method of Computing the MTF

Although the modulation indices may be measured directly by comparing system input to output modulation, Schroeder [17] derived the relationship between the source-to-receiver squared impulse response and the MTF. This relationship makes it possible to compute the MTF once having measured or predicted the squared impulse response. The Schroeder formula is:

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

where,

MTF is the modulation transfer function.

h(t) is the system impulse response.

F is the modulation frequency (Hz).

The modulation transfer function is therefore 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. Direct implementation of this equation, however, does not properly account for the effect of background noise on the STI. The STI algorithm specifies an input 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. However, if the speech signal to background noise ratio is known, its effect can be added after the MTF has been computed using a squared impulse response uncorrupted by noise [15].

1.4 Reported Accuracy of the STI Method

Steeneken and Houtgast [18] tested the accuracy of their method by measuring the STI in a large number of different conditions. These included conditions where bandpass limiting, background noise, peak clipping, automatic gain control and reverberation were responsible for degrading speech intelligibility. They scored speech intelligibility using trained listeners and Dutch monosyllabic nonsense words. A third-order regression curve was fit to the STI versus speech intelligibility data, and a standard deviation around this curve of $\sigma = 5.6\%$ was found. In practical terms this means that measurement of the STI results in a prediction of speech intelligibility which is within ± 5.6% of the actual speech intelligibility most of the time. When their analysis was confined to only those conditions where modulation reduction was due to distortions in the time domain, including background noise, reverberation, and automatic gain control, the standard deviation of the data about the regression curve was slightly higher at $\sigma = 5.8\%$.

2. COMPUTER IMPLEMENTATION OF THE 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 named "Modeler Design Program" [11]. Modeler uses a graphic user interface to facilitate rapid entry of room models and selection and placement of loudspeakers [20].

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 its octave-band Sabine absorption coefficients. Sound sources are represented by their full-space octave-band polar responses from 125 Hz to 4 kHz, and by their sensitivity and maximum power handling capability. Sound systems are modeled by specifying cluster locations, source aiming angles, electrical power requirements, and electronic time delays. An example of one of the rooms modeled in this study is shown in Fig. 3.

The program uses three algorithms to produce output relevant to the sound system designer. Direct field contributions at a point in the room are predicted by computing the inverse square loss from the source to the listener. Discrete early reflections are predicted using an image source method [13]. Late diffuse reverberation is predicted using the assumptions of statistical acoustics [14].

2.2 Hybrid Energy Decay Curve (HEDC): Rationale

As discussed in Section-1.3, the modulation transfer function and thus the STI can be computed using Eq. 1, which requires the squared impulse response as input. Prediction of the squared impulse response in a modeled room requires the prediction of an enormous number of reflections. On average, it can be shown [21] that the number of reflections arriving within a certain time is:

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

where,

N is the number of reflections arriving within t seconds.
c is the speed of sound in meters per second.
t is the time in seconds.
V is the volume of the room in cubic meters.

In a typical auditorium ($V = 6{,}000\ m^3$) 28 reflections arrive in the first one-tenth of a second, 3,522 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 enormous number of reflections which occur in typical rooms.

If Eq. 2 is differentiated with respect to time, the average number of reflections per unit time, or the reflection density, can be found:

$$\frac{dN}{dt} = \rho(t) = \frac{4\pi c^3 t^2}{V} \qquad (3)$$

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

This equation shows that the reflection density increases with time squared. As the reflection density increases, however, so does the diffuseness of the reverberant field. The room rapidly becomes filled with a great number of wavefronts whose behavior becomes more and more random. Under these conditions, it is possible to describe the acoustic behavior using statistical assumptions – to average the wavefronts instead of attempting to follow them independently. Accounting for the average behavior of diffuse 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 the discrete early reflections; these are the arrivals which are uniquely characteristic of the sound system design (speaker types, speaker aiming angles, power levels, and delay) and the room (geometry and specific distribution of absorption). A statistical model of reverberation is then used to account for the late diffuse reverberation. The result is a representation of energy transmission which exploits the sophistication of the computer room model when it is still computationally efficient to do so, and then switches to a much simpler model to account for the remaining late statistical reverberation. The choice of when to switch from one model to the other depends on the specifics of the sound system design and room model. A graphical representation of the HEDC, consisting of discrete early arrivals and late diffuse reverberation is shown in Fig. 4.

3. EXPERIMENT TO TEST THE ACCURACY OF NEW STI METHOD

3.1 Test Rooms, Sound Sources, and Listener Positions

Ten rooms, three sound sources, and two listener positions per room comprised a data base of fifty different conditions for speech intelligibility. The details of these conditions are extensively described elsewhere [10]. The rooms ranged in size, architectural complexity and reverberation characteristics. The sources were chosen for their wide range of polar responses, and listener positions were chosen to represent positions both near and far from the sources. These conditions are summarized in the tables below.

Table-1. Room Parameters

Name	T60	Function
Berklee Performance Center	0.9[2]	Music
Coolidge Corner Movie House	1.0	Cinema
Huntington Theater	1.1	Drama
Saint Bridget Church	2.0	Religious
Nevins Hall	3.5	Multi-function
Jordan Hall	2.2	Music
Mechanics Hall	2.2	Music
South End Cathedral	3.3	Religious
Cyclorama	3.5	Multi-function
MIT Indoor Track	4.6	Athletics / P.A.

Table-2. Loudspeakers

Name	Type	Directivity [3]
Soundsphere 2212-1	Omni-radiator	1.1
Bose 802-II	Eight driver array	7.3
Electro Voice HR6040A	Constant directivity horn (with TL806AX)	17.7

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

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

Table-3. Listener Positions

Name	Relationship to Source	Position in Room
Near position	On axis ±7.5°	1/3 of room length
Far position	On axis ±7.5°	Rear of room

3.2 Subject-Based Testing

Subject-based intelligibility tests were conducted for every combination of room, sound source, and listener location, the exact details of which are described elsewhere [10]. The tests were administered according to the American National Standards Institute "Standard for Measuring Monosyllabic Speech Intelligibility" (ANSI S3.2-1971). Subject-based intelligibility scores are denoted *%PB-ansi* in this study. The total number of words presented for each room, source, and listener position combination ranged from 2,000 to 2,800 words. Mean scores for the fifty conditions are shown in the Appendix.

3.3 In-Room Measurements

For each room, source, and listener position combination, system impulse responses were measured and recorded. The STI for each of these measured impulses was computed by applying Eq. 1 and the MTF-to-STI conversion mathematics specified by Houtgast and Steeneken [15]. These values are denoted *STI-measured*, and are tabulated in the Appendix.

3.4 Room Modeling and STI Prediction

Room models were created in the computer program for each of the ten rooms. Materials were chosen from a standard list of materials [14], 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 then used to compute an octave-band modulation transfer function. (The 8 kHz octave band was simply a copy of the 4 kHz band, but was subsequently weighted according to the algorithm specified by Steeneken and Houtgast.) Finally, the modulation transfer functions were condensed to the Speech Transmission Index; these values are denoted *STI-predicted*, and are tabulated the Appendix.

4. RESULTS

4.1 Relationship Between STI-measured and %PB-ansi

A third order polynomial regression curve was computed for the STI-measured versus %PB-ansi data. The regression curve and data are shown in Fig. 5. The standard deviation of the data about the regression curve is 5.2%, which is similar to the 5.8% value reported by Steeneken and Houtgast. The regression curve equation is:

$$\%PB\text{-}ansi = 788.26STI^3 - 1643.9STI^2 + 1179.3STI - 196.3 \quad (4)$$

4.2 Relationship Between STI-predicted and %PB-ansi

The STI-predicted values were converted to %PB-ansi scores using Eq. 4; the data are shown in Fig. 6. The standard deviation of this data about the regression curve is 5.4%. Thus the overall error in predicting speech intelligibility using STI-predicted values is essentially equivalent to the error using STI-measured 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.

4.3 Relationship Between STI-predicted and STI-measured

While the preceding result shows that speech intelligibility can be predicted with essentially the same accuracy using STI-predicted values as with STI-measured, it is also of interest to study the direct relationship between the STI-predicted and STI-measured values. A scatter plot showing this relationship is shown in Fig. 7. The correlation coefficient for the data is r = 0.81, which is considered good to very good. Thus the HEDC can be considered as a reasonable substitute for the actual squared impulse response in this application.

5. DISCUSSION

5.1 Overall Accuracy of the STI Method

Results show that predictions of speech intelligibility using the new computer-based implementation of the STI method are essentially as good as those based on measured STI values. This is an important step in providing the sound system designer with a tool for predicting speech intelligibility in unbuilt or inaccessible rooms. However, it is important to note the significance of a standard error of estimation of 5 to 6%. The

Houtgast and Steeneken data [18], and the measured and predicted STI data from this study each show standard deviations in this range. A standard deviation of 5 to 6% means that the subject-based intelligibility scores will be within one standard deviation of the predicted intelligibility most of the time. Therefore an error of ± 5 to 6% intelligibility must be included in interpreting predictions based on the Speech Transmission Index.

The 5 to 6% error inherent in the STI method, while no greater than some other published methods and much less than others [10], may be reduced in future studies through refinement. 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 primarily responsible for producing the consonant sounds.

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 of predicting early reflections and statistical acoustics theory to predict late diffuse reverberation. Both of these models have known limitations. One is the assumption that a room boundary can be approximated by a flat plane with a single octave-band absorption coefficient. Real surfaces can both reflect and scatter sound waves, and have absorption properties which are a function of incident wave angle. In addition, the prediction of reverberant decay rates using statistical assumptions is, by the nature of statistics, only an estimate. Therefore rooms with large complicated scattering surfaces or unusual reverberation characteristics are likely to result in higher errors.

The fifty conditions used in this study each represent conditions where reverberation is responsible for degrading speech intelligibility. Background noise was minimized as a factor by guaranteeing in each case that the speech signal-to-background noise ratio exceeded 15 dB. (The emphasis on the effect of reverberation was intentional in order to test the suitability of the Hybrid Energy Decay Curve.) Thus the effect of background noise was not explicitly tested. However, this effect has been extensively studied by Houtgast and Steeneken [15, 18, 19] and others [8, 9, 22].

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

5.3 Relationship Between STI and Subject-Based Tests

The relationship established in this study between the STI and %PB-ansi (Eq. 4) is unique. Other studies have used different subject-based tests, methods of training subjects, or test conditions, and have therefore arrived at different regression curves relating STI to speech intelligibility. The details of these relationships including the regression curves, are given in the Appendix.

6. CONCLUSION

The Speech Transmission Index method of estimating speech intelligibility has been implemented in a microcomputer-based program for predicting sound system performance. The STI implementation requires no in-room measurements and is thus suitable for unbuilt or inaccessible sound systems and rooms. The new technique relies on a computationally efficient method of representing the transmission of energy 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 diffuse reverberation.

Results of an experiment designed to test the accuracy of this new implementation show that essentially 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 ± 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 shows that the HEDC is a good substitute both for the actual source-to-receiver energy vs. time response and other computationally more intensive representations of this response.

The overall accuracy of the STI method was shown to be important in terms of interpreting predictions of speech intelligibility. While predictions based on these STI values are as good, and is some cases much better than other methods, sound system designers should be mindful that the intelligibility estimates 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 new STI method is implemented since all of the details necessary to generate the HEDC have not been presented. 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," Inst. for Perception TNO, Report #IZF 1982-29 (1982).

[2] Brüel and Kjær Technical Review, "RASTI," No. 3, (1985).

[3] D. Keele Jr., "Evaluation of Room Speech Transmission Index and Modulation Transfer Function by the use of Time Delay Spectrometry," *Proc. of the AES 6'th Intl. 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 *Natl. Elect. Conf.*, Chicago, Oct. (1981).

[6] H. Kuttruff, *Room Acoustics*, 2nd ed., Appl. Science Publ.: London (1979), p. 180. (See for formulas other than the STI.)

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

[8] B. Anderson and J. Kalb, "English Verification of the STI method for Estimating Speech Intelligibility of a Communications Channel," *J. Acous. Soc. Am.*, Vol. 81, No. 6, (1987).

[9] J. Bradley, "Predictors of Speech Intelligibility in Rooms," *J. Acous. Soc. of Am.*, Vol. 80, No. 3, (1986).

[10] K. Jacob, "Correlation of Speech Intelligibility Tests in Reverberant Rooms with Three Predictive Algorithms," *J. Aud. Eng. Soc.*, Vol. 37, No. 12, (1989).

[11] The program, called Modeler® Design Program, is a registered trademark of Bose Corporation.

[12] Hybrid Energy Decay™ Curve and HEDC™ are trademarks of Bose Corporation.

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

[14] L. Beranek, *Acoustics*, Amer. Inst. Phys. for the Acoust. Soc. Am., New York, pp. 304-307, (1986).

[15] T. Houtgast, 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," *Proc. of the Reg. Conf. of the Aud. Eng. Soc.*, Tokyo Japan, June (1989).

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

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

[19] 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).

[20] T. Birkle and K. Jacob, "Modeler Design Program v3.0," *Sound and Video Contractor*, Vol. 7, No. 8, (1989).

[21] H. Kuttruff, op. cit., p. 87.

[22] K. Kryter and E. Whitman, "Some Comparisons Between Rhyme and PB-word Intelligibility Tests," *J. Acous. Soc. Am.*, Vol. 37, p 1146 (1965).

[23] T. Houtgast, H. Steeneken, "Evaluation of Speech Transmission Channels by Using Artificial Signals" *Acustica*, Vol. 25, (1971).

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

8. APPENDIX

8.1 Relationship Between STI, %PB-ansi and other Subject-Based Tests

In this study, phonetically balanced (PB) monosyllabic English words were used as specified by an ANSI standard (S3.2-1971). The relationship between the STI and %PB-ansi, shown as the regression curve of Figs. 5 and 6, is applicable to the ANSI test only. Other studies have used other subject-based tests, and the relationships established between the STI and these tests are different. Because of these differences, the subject-based test must be specified in converting the STI to subject-based speech intelligibility scores; the method of testing, the type of words, and the language all must be included since each will affect the relationship to the STI. The relationship between the STI and various subject-based tests determined by other studies are discussed here.

Houtgast and Steeneken used phonetically balanced Dutch consonant-vowel-consonant (CVC) nonsense words [19]. Bradley [9] measured intelligibility using the Fairbanks Rhyme Test, which is a multiple choice test where the words in each set rhyme. Anderson and Kalb [8] 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 [22]. The relationship between the STI and the various methods of measuring speech intelligibility established by these studies is shown in Fig. 8.

Inspection of these regression curves 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. Scores from this study are slightly lower than those obtained by Houtgast and Steeneken. 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 [23] for each of their conditions. This represents a factor of four difference in testing time.

Fig. 8 also shows that subjects score much higher on the two rhyme tests than the other measures of speech intelligibility. This is expected based simply on the fact that the rhyme tests are multiple choice (closed set) tests and choices differed only in one consonant. CVC words contain by definition two consonants, and PB words average almost two per word.

The Anderson and Kalb curve shows that their subjects scored significantly higher in intelligibility tests than those in this study or those of the Houtgast and Steeneken study. The most obvious explanation is the fact that Anderson and Kalb "thoroughly familiarized" their subjects with the words before beginning testing. Thus, in a sense, their subjects were choosing from a closed set, and may have memorized the words. Under these conditions, an increase in the intelligibility scores would be expected. Another possible explanation is based on the fact 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 artificial conditions for which the STI measure was not 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 inventors of the STI method was shown to be similar to the American Standards 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 work has been conducted in this area [24].

8.2 STI-measured, STI-predicted, and Subjective Data

Exact experimental details of subject-based testing can be found in [10]. In the table below: *STI-meas.* refers to in-room measurements of the Speech Transmission Index, *STI-pred.* refers to the predicted STI (based on the Hybrid Energy Decay Curve) from the computer program, and *%PB-ansi* refers to the mean score on intelligibility word lists.

Condition			**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
Mechanic's	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

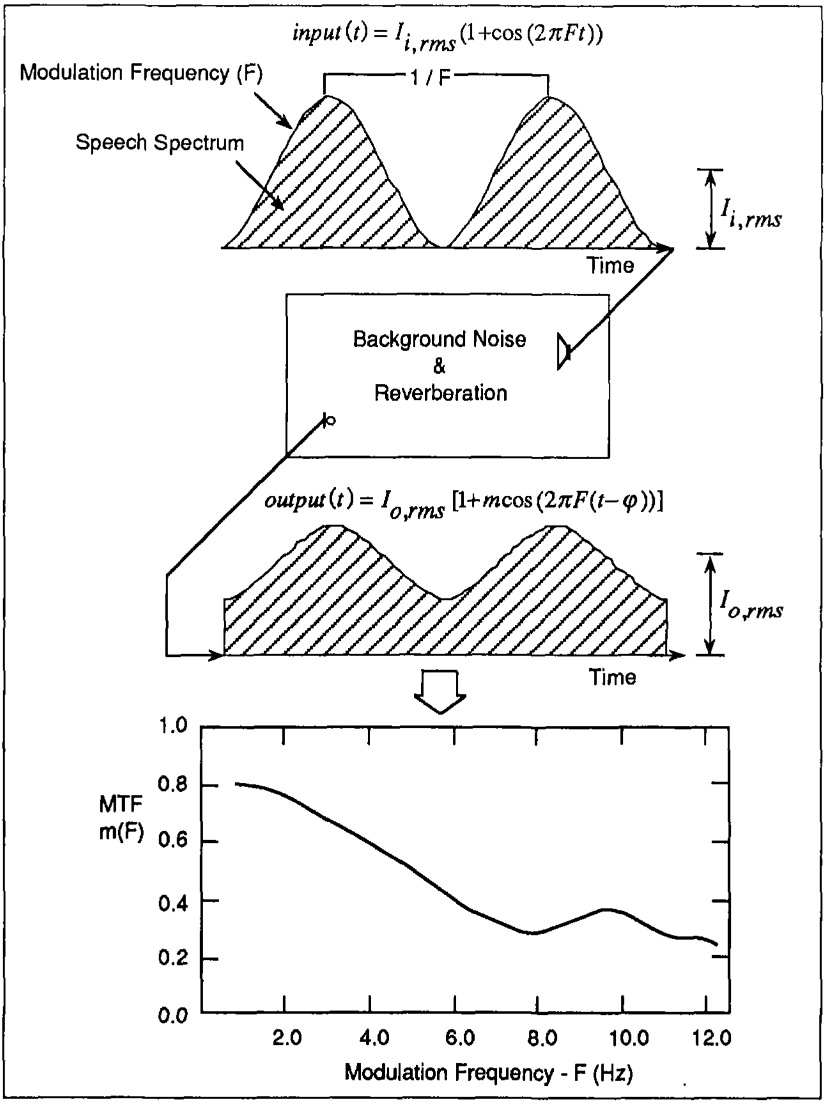

Fig. 1. The STI method uses an artificial speech signal modeled after the behavior of actual speech. This signal is an amplitude-modulated speech spectrum. In this figure, the blurring effect background noise and reverberation have on the input waveform is shown. The modulation index (m) is a measure of the preservation of the original modulation at the input of the system, and the Modulation Transfer Function (MTF) is the modulation index as a function of modulation frequency (F). (Figure attributable to Houtgast and Steeneken [15].)

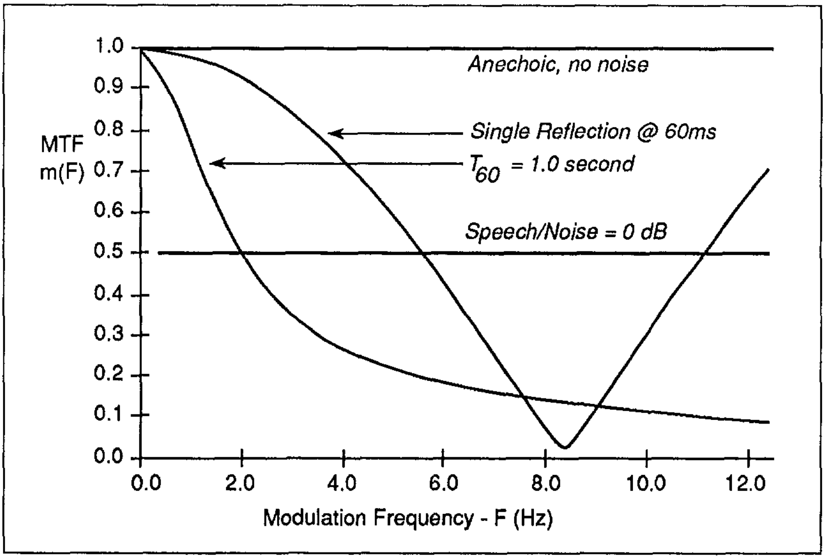

Fig. 2. The Modulation Transfer Function (MTF) is a good diagnostic tool since its shape reflects the acoustic conditions responsible for reducing speech intelligibility. Several examples are shown in this figure. Perfect preservation of the original modulation results in no reduction of the modulation index. Reduction due to background noise alone results in an overall reduction of the modulation index. Reduction due to pure exponential reverberation results in a monotonically decreasing modulation index with modulation frequency. Reduction due to a single reflection causes a deep notch in the MTF.

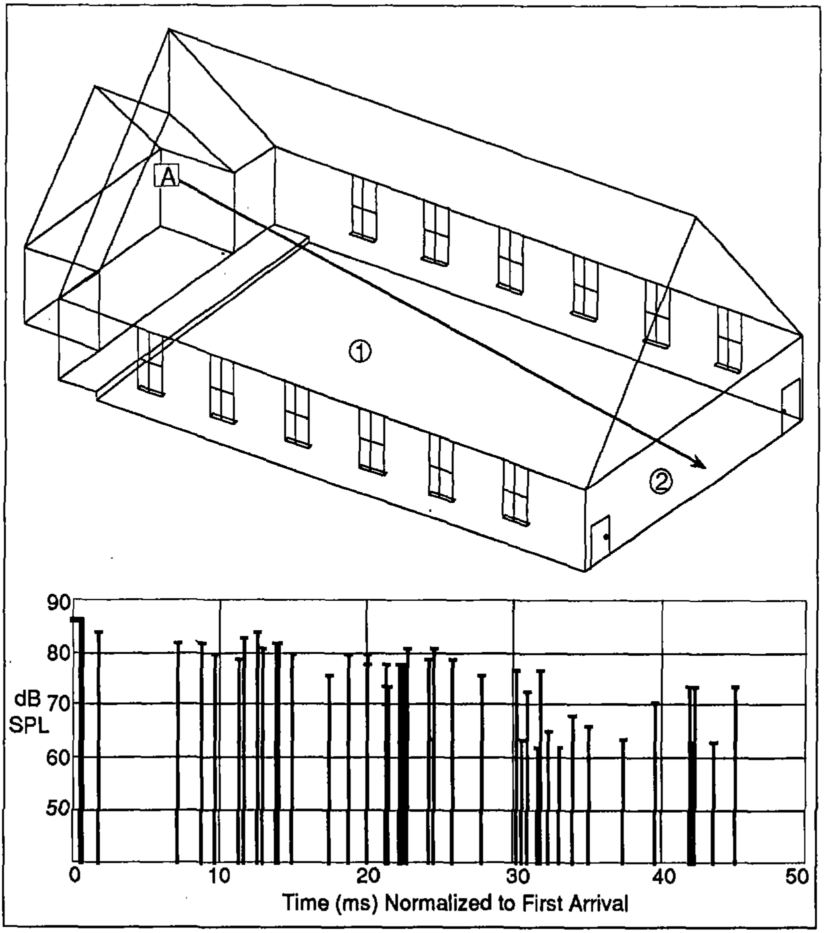

Fig. 3. A medium-sized church as modeled in the computer program is shown. Rooms are constructed as a series of adjacent planes, each of which is assigned a surface material. Sound sources are represented by their full-space polar responses from 125 Hz to 4 kHz. The early-energy vs. time response (consisting of direct field and early reflections) is shown for the position at the rear of the room.

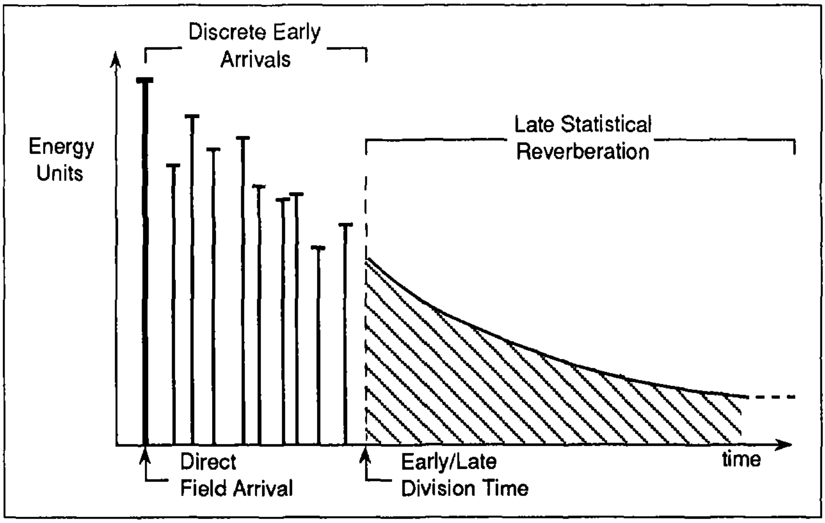

Fig. 4. A graphical representation of the Hybrid Energy Decay Curve (HEDC) is shown. The HEDC consists of an early part – composed of direct field and early reflections predicted using an image source method – and a late part representing diffuse reverberation and predicted using classical reverberation theory.

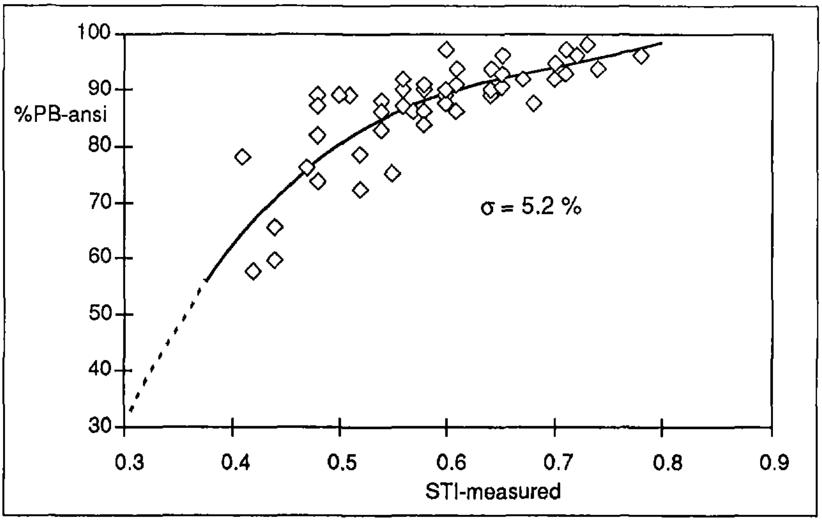

Fig. 5. A scatter plot of STI-measured (from the ten-room data base) versus speech intelligibility from the ANSI subject-based test is shown, along with the third-order regression curve. The average distance of the points from the regression curve, or the error in predicting intelligibility from the measured STI, is 5.2%.

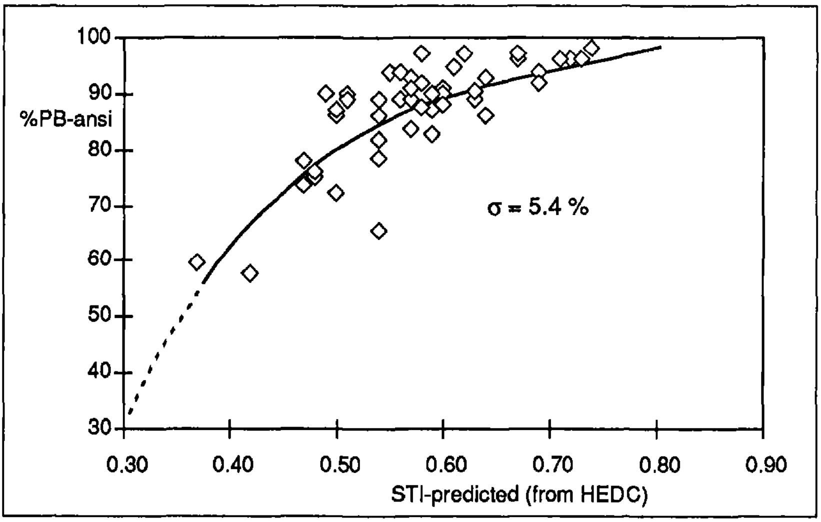

Fig. 6. STI-predicted (from the computer-generated HEDC) versus speech intelligibility from the ANSI test is shown, along with the third-order regression curve of Fig. 5. Notice that the closeness of data to the regression curve is essentially the same as the data of Fig.-5.

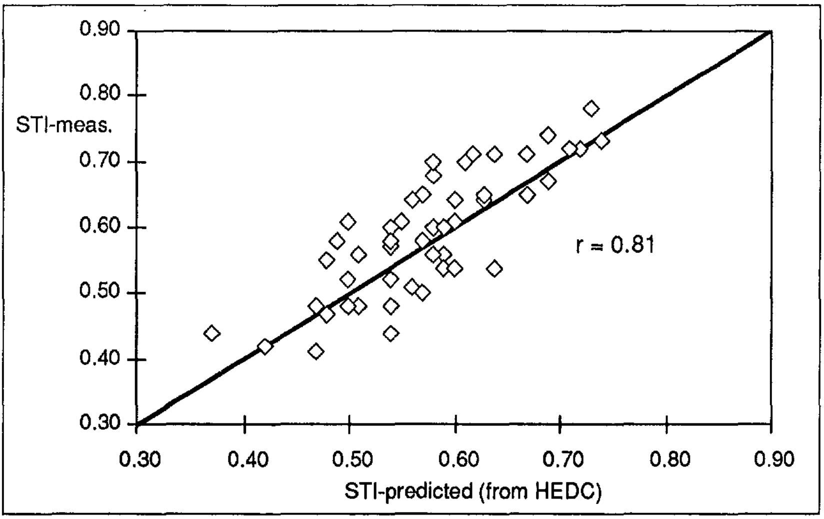

Fig. 7. A scatter plot of STI-predicted (from the computer-generated HEDC) versus STI-measured is shown. The straight line represents perfect correlation. The correlation coefficient of r = 0.81 is good to very good.

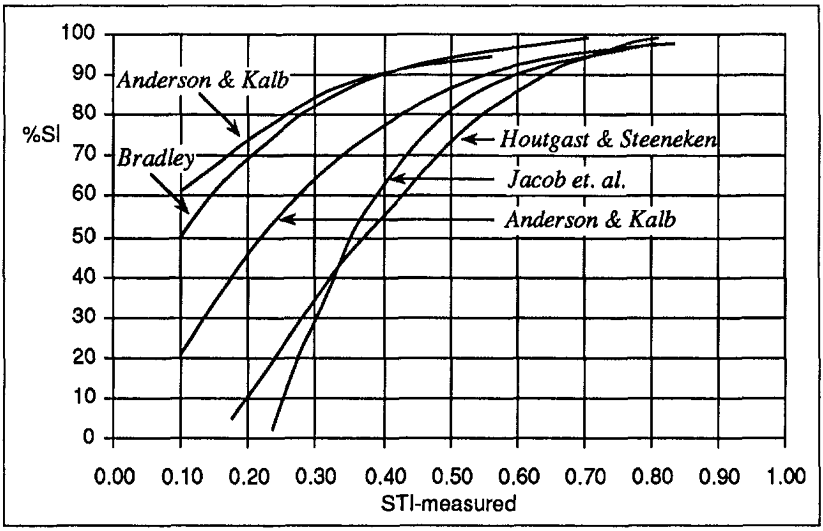

Fig. 8. The relationships between the Speech Transmission Index (STI) and various subject-based intelligibility tests used in this and other studies are shown. The differences are primarily the result of different test types. The top two curves are from rhyme tests. The two middle curves are from tests using monsyllabic English words. The lowest curve (from Houtgast and Steeneken) is from a test using monosyllabic nonsense words.