Evaluation of Room Speech Transmission Index and Modulation Transfer Function by the Use of Time Delay Spectrometry

D. (Don) B. KEELE, JR.

Techron, Div. Crown International, Inc., Elkhart, Indiana, 46517, USA

The literature shows that the modulation transfer function (MTF) and speech transmission index (STI) can be computed from the squared impulse response of a linear passive system. This paper describes an extension of this method to measurements of systems using time delay spectrometry (TDS). The new method makes use of both the real and imaginary parts of the complex analytic impulse response of the system (the energy-time response). This allows more accurate determination of STI and MTF because the calculations are based on measurements that more closely follow the true energy decay in the room.

0. INTRODUCTION

M. R. Schroeder has shown an alternate method for computing the modulation transfer function (MTF) of a linear passive system [1]. This method shows how the MTF can be calculated from the impulse response of a system without having to directly measure the modulation transfer characteristics at each individual modulation frequency.

This paper briefly describes this method and shows how the method can be used with a time delay spectrometry based analysis system to evaluate the speech transmission index. The software runs on the Techron TEF System 12 analyzer and allows quick and easy field measurements of speech intelligibility.

1. SPEECH INTELLIGIBILITY MEASUREMENTS

Several methods exist for predicting speech intelligibility in rooms that are based on making measurements of parameters such as speech signal-to-noise ratios and room acoustics properties. These methods use these acoustics measures to predict speech intelligibility using various algorithms and procedures described in the literature.

Bradley [2] gives a good overview of the different methods which include the work of Haas [3], Lochner and Burger [4], Peutz [5], Houtgast and Steeneken [6], and Bradley [7-9]. The Syn Aud Con Tech Topics of Davis are also a rich source of intelligibility measurement information [10].

2. MODULATION TRANSFER FUNCTION

The modulation transfer function is a measure of how well the amplitude modulation (variation of intensity with time) of a signal is preserved when the signal is sent from one point to another in a particular transmission chain.

Research has shown that the a good portion of the intelligence in human speech is contained in the modulation of the speech waveform [11] {Authors comment: debate has come up recently about this point, see comments by Davis [10]}. Preservation of the speech modulation patterns is important to maintain high intelligibility. Noise, echoes and reverberation are found to decrease the effective modulation of the speech waveform and hence impair intelligibility.

The MTF is usually plotted as a function of modulation frequency with the values ranging from 0 to 1. This range represents no transmission of modulation (zero) to perfect transmission of modulation (unity). Typical speech modulation frequencies are in the range of 0.5 to 16 Hz. Music modulation frequencies can range up to much higher values [12]

3. STI AND RASTI

Houtgast and Steeneken [13] describe a specific method for evaluating speech intelligibility using a direct measurement of the MTF based on sinewave amplitude modulation of a band-limited random noise carrier signal. Their method describes a series of modulation transfer measurements with carrier frequencies ranging over 125 Hz to 8 kHz at octave center frequencies. Typical human voice modulation frequencies of 0.5 to 16 Hz, at one-third octave frequency intervals, are used for modulation.

These measurements result in a data set containing 98 modulation reduction indexes which are in turn converted into a single index called the Speech Transmission Index (STI). The STI value is a single number that indicates the effect of a transmission system on speech intelligibility.

A shortened version of this procedure called RASTI (Rapid Speech Transmission Index) is also described which only requires the measurement of 9 modulation reduction factors distributed between the 500 Hz and 2 kHz octaves. Complete information on this technique can be found in B&K's Technical Review [14] and also in [15]. The B&K Sound Transmission Meter Type 3361 uses the RASTI technique for its measurements.

4. ALTERNATE MTF COMPUTATION METHOD

M. R. Schroeder has shown an alternate method for computing the MTF [1]. This method computes the MTF from the impulse response of a linear passive system without having to directly measure the modulation transfer characteristics at each individual modulation frequency. The method derives the MTF by calculating the normalized frequency spectrum of the squared impulse response (energy decay) of the system.

Schroeder shows that the modulation transfer function (MTF) can be calculated from the impulse response of a linear passive system:

$$\mathrm{m}(\omega) = \frac{\int_0^\infty h^2(t)e^{-i\omega t}dt}{\int_0^\infty h^2(t)\,dt} \qquad (1)$$

where $\mathbf{m}(\omega)$ is the complex (having magnitude and phase) modulation transfer function, $h(t)$ is the impulse response of the system, i is the complex imaginary unit operator (square root of minus one), and ω is the frequency in radians per second (2 π times the frequency).

The numerator of the equation is recognized as being the Fourier transform or frequency spectrum of the square of the impulse response and the denominator is just the total energy of the impulse response. The ordinary or traditional MTF is the absolute value of the complex MTF, $|\mathbf{m}(\omega)|$ and ranges over 0 to 1.

5. STI MEASUREMENTS USING FFT

Hojberg [16] indicates how this computation method can be applied to STI intelligibility measurements using a dual-channel FFT spectrum analyzer. The impulse response of a system is first measured using the dual-channel analyzer and then the data is converted to MTF and STI intelligibility data by post processing with a computer. Hojberg discloses that the measurement and computation time for a complete STI determination takes about 15 minutes using his method

6. STI MEASUREMENTS USING TDS

TDS can also be used measure the impulse response of a system [17-18]. This allows the MTF and STI data to be derived from TDS measurements rather than FFT measurements.

The use of TDS to measure the system impulse response has several advantages over the FFT based measurement technique [19]. These including: fast measuring time, superior noise rejection (when required), low crest factor test signal and more freedom to tailor the frequency content of the test signal to the desired operating range of the system being tested.

Rather than using the squared impulse response data alone to calculate the MTF, the squared TDS Energy Time Curve (ETC) data is used to calculate the MTF. The squared ETC data is recognized to be precisely the energy of the envelope of the systems impulse response with no approximations. As will be shown in a future technical paper, this squared ETC data is found to better represent the actual system energy decay response and hence yield a more accurate MTF calculation.

The impulse response can be measured with TDS methods using the ETC or Energy-Time Curve data. The MTF and STI can then be evaluated yielding the speech intelligibility prediction.

Because TDS naturally yields the complete complex analytic impulse response $\mathbf{h}(t)$ [18], which includes the impulse response (real part, $f(t)$) and the doublet response (imaginary part, $g(t)$):

$$\mathbf{h}(t) = f(t) + ig(t), \qquad (2)$$

this knowledge can be used to calculate a more accurate MTF. Note the bold face which denotes a complex quantity.

In this case, a modified version of Schroeder's equation is used with the squared impulse response $\left(h^2(t)\right)$, replaced with the square of the absolute value of the analytic impulse response $\left(|h(t)|^2 = f^2(t) + g^2(t)\right)$

$$m(\omega) = \frac{\int_0^\infty |h(t)|^2 e^{-i\omega t} dt}{\int_0^\infty |h(t)|^2 dt} \tag{3}$$

7. NEW TEF SOFTWARE

A new software package for the TEF System 12 Analyzer, has been developed, that does speech intelligibility measurements based on the previous concepts. The software can perform a complete STI measurement over the range of 125 Hz to 8 kHz in 2 minutes. It can also measure a TDS equivalent of the RASTI method by testing only at the 500 Hz and 2 kHz octaves. Limiting the test to two octaves yields a quicker test time of about 30 seconds.

The full STI test is accomplished by measuring seven individual one second time span ETC's at each of the octave center frequencies between 125 Hz and 8 kHz. The background noise in each octave band is evaluated in a separate measurement. After each ETC test, the MTF is calculated and the STI in each octave band is computed. The software allows for separate direct entry of octave noise levels to check the effect of noise on STI or RASTI values. The TEF test generator level at each octave band is adjusted to match the average spectral content of speech. At the conclusion of the test, the overall STI value is computed by taking a weighted average of the individual octave band STI values.

The weights used for the full STI averaging follow the importance of each of the individual octave bands to speech intelligibility. These weights are based on the work of French and Steinberg [20] and are found to emphasize the 1 to 4 kHz octave bands. Note that the weights used here *do not* agree with the roughly equal octave weighting of Steeneken and Houtgast [14] but closely agree with the more recent work of Humes et al [21] in their hybrid index mSTI (see Fig. 1).

The TEF-RASTI equivalent measurement gives equal weights to the 500 Hz and 2 kHz octave bands which is in agreement with the weighting used in B&K's RASTI technique.

The TEF-RASTI measurement method used here also makes use of all the one-third-octave modulation data from 0.5 to 12.5 Hz for both octave bands. This is in contrast with the conventional RASTI technique, as implemented by the B&K unit, which uses only the modulation data from 4 one-third-octaves bands at 500 Hz and 5 bands at the 2 kHz octave centers. This sparse spacing of modulation data means that in some situations incorrect intelligibility numbers will be predicted.

The software provides for complete data storage and retrieval of all raw ETC test data. This allows freedom in future post processing of the data to yield new information.

8. TEF-STI DATA AND DISPLAYS

The TEF-STI software can display several different forms of tabular and graphical data both on the screen and on printouts. All displays show the overall STI value along with the subjective qualification of the speech intelligibility (bad, poor, fair, good, excellent, etc.).

The tabular data includes octave specific and overall data for: STI, equivalent early decay time (Early RT_{60}), and equivalent signal to noise ratio (S/N Ratio). Refer to [13-14] for explanations of these latter two parameters.

The graphical data that can be selected includes plots of the individual octave band ETC's, MTF's, and a display of the STI values, S/N ratios, and early RT_{60} times versus frequency.

Observe that all modulation transfer function data is displayed over the rather wide frequency range of 0 to 155 Hz. This allows assessment of a particular path's transmission characteristics not only for speech modulation frequencies (0.5 to 16 Hz) but for other sources, such as music, that have higher modulation frequencies.

9. EXPERIMENTAL MEASUREMENTS

Several experimental measurements were made with the software to validate the computation methods. Measurements were made on a digital reverberation simulator, a "loop-back" connection with added noise, and in a large factory acoustic space. No direct comparisons were made with actual subjective intelligibility tests. This will be done at a later date.

9.1. Reverberation Simulator

An Alesis Microverb reverberation simulator was used as an experimental unit for measurements. The measurement was "noise free"; no additional noise was added to the signal. Figs. 2 and 3 shows the results of TEF-RASTI and ETC measurements with the simulator set to the number 5 small room configuration. This setting provides an approximate reverberation time (RT_{60}) of about 1.8 seconds with no initial time delay gap.

The TEF-RASTI intelligibility measurement (Fig. 2) yielded a RASTI number of 0.54 with a "fair" subjective evaluation. The 2 kHz octave data indicates an equivalent signal-to-noise ratio of 1.3 dB and an equivalent early reverberation time of 1.27 seconds.

Fig. 3 shows the 2 kHz energy decay (ETC) data that was used in the RASTI determination. Shown on the figure is a 1.27 second decay line fitted to the early decay of the response (first 5 dB of decay). Note that in this case the predicted intelligibility is better because the initial 5 to 6 dB decay rate (1.27 second equivalent early-decay RT_{60}) is faster than the long-term decay rate (1.8 second equivalent RT_{60}).

9.2. Loop-back Transmission Path With Noise

To check the effect of noise on an otherwise perfect transmission channel, 2 kHz one-third-octave bandwidth, pseudo-random noise was added to a loop-back transmission path. A zero dB signal-to-noise ratio situation was created by setting the RMS level of the noise to the TEF's sinewave generator level. Fig. 4 shows the results of a TEF-RASTI test performed on this transmission setup.

The 2 kHz data shows a STI value of 0.49 and an equivalent S/N ratio of -0.2 dB. Theory predicts a STI value of 0.5 and an equivalent S/N ratio of 0 dB. Note that the 2 kHz MTF data is reduced by approximately one-half at all modulation frequencies, as expected.

The 500 Hz data shows how a perfect transmission path measures (STI = 1.00). The 500 Hz MTF graph (bottom) shows moderate rolloff above 80 Hz modulation frequencies. This is due to the raised cosine weighting function used to derive the ETC data.

9.3. Large Acoustic Space

TEF-RASTI and TEF-STI measurements were done in a large empty factory space of about 21,000 cu m (750,000 cu ft). Two sets of measurements were made on a small stand-mounted low-directivity two-way loudspeaker monitor. The first measurement was taken close to the monitor at a distance of 3 m (10 ft) and the second at 34 m (110 ft). The near measurement would presumable yield high intelligibility while the far measurement would yield poorer intelligibility.

Fig. 5 shows the results of these intelligibility tests. Fig. 5a shows the TEF-RASTI results, while Fig. 5b shows the TEF full STI results. As expected, the far measurements yielded lower intelligibility.

Note the comb filtering in the 500 Hz MTF data in the FAR location that resulted from a high level close-in reflection in the 500 Hz ETC data. Note also the difference in the overall STI values at the farthest location between the TEF-RASTI (= 0.61, "Good intelligibility") and TEF-STI (=0.54, "Fair intelligibility").

The TEF-STI measurement yielded a lower intelligibility prediction because it gives greater weight to the 1, 2, and 4 kHz octaves, which happened to have lower individual STI values. In contrast, the TEF-RASTI measurement, which only measures the 500 Hz and 2 kHz octave bands, yields a higher value because the 500 Hz band happened to have a higher STI value of 0.78.

In most cases, the TEF-STI intelligibility prediction is more accurate than the TEF-RASTI prediction because more data is measured and analyzed. The TEF-STI prediction is always more accurate if there are large data differences between octave frequency bands.

10. SUMMARY

This paper has briefly reviewed some of the features and development details of a new TEF System 12 software package that allows rapid and easy objective measurements of speech intelligibility of sound systems, acoustic spaces, and electronic transmission links. The software is based on the modulation transfer function and the speech transmission index and is measured using time delay spectrometry techniques.

11. REFERENCE LIST

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

[2] J. S. Bradley, "Review of Methods of Measuring Rooms for Speech," Presented at the Acoust. Soc. Am. meeting, Indianapolis, Indiana (May 1987)

[3] H. Haas, "The Influence of a single Echo on the Audibility of Speech," Reprinted in *J. Aud. Eng. Soc.*, Vol. 20, No. 2, pp 146-159 (1979)

[4] J. P. A. Lochner and J. F. Burger, "The Influence of Reflections on Auditorium Acoustics," *J. Sound Vib.*, Vol. 1, pp 426-454 (1964).

[5] V. M. A. Peutz, "Articulation Loss of Consonants as a Criterion for Speech Transmission in a Room," *J. Aud. Eng. Soc.*, Vol. 19, No. 11, pp 915-919 (1971)

[6] T. Houtgast and H. J. M. Steeneken, "A Review of the MTF Concept in Room Acoustics and its use for Estimating Speech Intelligibility in Auditoria," *J. Acoust. Soc. Am.* 77, No. 3, pp 1069-1077 (1985)

[7] J. S. Bradley, "Predictors of Speech Intelligibility in Rooms," *J. Acoust. Soc. Am.* 80, No. 3, pp 837-845 (1986)

[8] J. S. Bradley, "Uniform Derivation of Optimum Conditions for Speech in Rooms," National Research Council Canada, *Building Research Note BRN 239*

[9] J. S. Bradley, "Auditorium Acoustics Measures from Pistol Shots," J. *Acoust. Soc. Am.* 80, No. 1, pp 199-205 (1986)

[10] D. Davis and C. Davis, "More on Intelligibility," *Syn Aud Con Tech Topics*, Vol. 14, No. 8, (1987)

[11] T. Houtgast and H. J. M. Steeneken, "The Modulation Transfer Function in Room Acoustics as a Predictor of Speech Intelligibility," *Acustica*, Vol. 28, p 66 (1973)

[12] J. D. Polack, H. Alrutz and M. R. Schroeder, "The Modulation Transfer Function of Music Signals and its Applications to Reverberation Measurements," *Acustica*, Vol. 54, pp 257-265 (1984)

[13] T. Houtgast and H. J. M. Steeneken, "The Modulation Transfer Function in Room Acoustics," *Bruel & Kjaer Technical Review No. 3*, pp 3-12 (1985)

[14] H. J. M. Steeneken and T. Houtgast, "RASTI: A Tool for Evaluating Auditoria," *Bruel & Kjaer Technical Review No. 3*, pp 13-30 (1985)

[15] Draft IEC Publication 268: Sound System Equipment - Part 16, "Report on the RASTI Method for the Objective Rating of Speech Intelligibility in Auditoria"

[16] K. Hojberg, "STI Measurements Using a Dual Channel Analyzer," Presented at the 81st Convention of the Aud. Eng. Soc., Nov. 1986, Preprint No. 2397 (E-7)

[17] R. C. Heyser, "Acoustical Measurements by Time Delay Spectrometry," *J. Aud. Eng. Soc.*, Vol. 15, p 370 (1967)

[18] R. C. Heyser, "Determination of Loudspeaker Signal Arrival Times, Part I," *J. Aud. Eng. Soc.*, Vol. 19, No. 10, pp 734-743 (1967)

[19] D. B. Keele, Jr., "TDS: Application of Technology," *Sound and Video Contractor*, Sept. (1985)

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

[21] L. E. Humes, S. Boney and F. Loven, "Further Validation of the Speech Transmission Index (STI)," *J. Speech and Hearing Research*, Vol. 30, pp 403-410 (Sept. 1987)

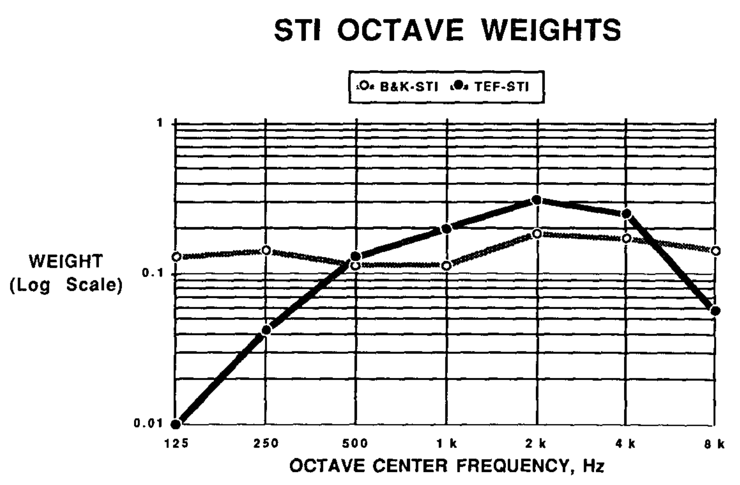


Fig. 1. Comparison between the octave band STI weightings used in the conventional (B&K) STI test and the TEF-STI test. Note that the conventional values gives roughly equal weighting to each octave while the TEF-STI weights follow the importance of each of the octave bands to speech intelligibility. The TEF-STI weighting emphasizes the 1 kHz to 4 kHz bands and deemphasises the frequency extremes. The following table lists the numeric weight values:

Center Freq., Hz =	125	250	500	1 k	2 k	4 k	8 k	Sum
B&K-STI Weight =	0.129	0.143	0.114	0.114	0.186	0.171	0.143	1.000
TEF-STI Weight =	0.010	0.042	0.129	0.200	0.312	0.250	0.057	1.000

REVERBERATION SIMULATOR

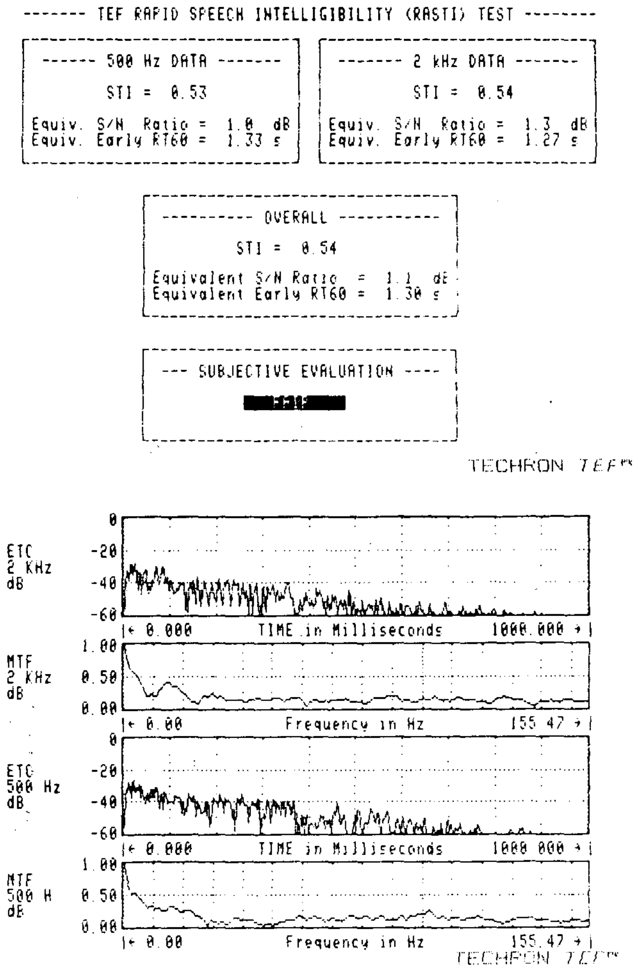


Fig. 2. TEF-RASTI intelligibility measurement on Alesis Microverb reverberation simulator. The simulator was set to the number 5 small room configuration. No external noise was added. The test resulted in a RASTI number of 0.54 with a "Fair" subjective intelligibility evaluation. The 2 kHz data shows a measured equivalent early decay time of 1.27 seconds. Tabular TEF-RASTI data is shown in the top half of the figure while the 500 Hz and 2 kHz band modulation transfer function (MTF) data and energy decay versus time (ETC) data is shown in the bottom half.

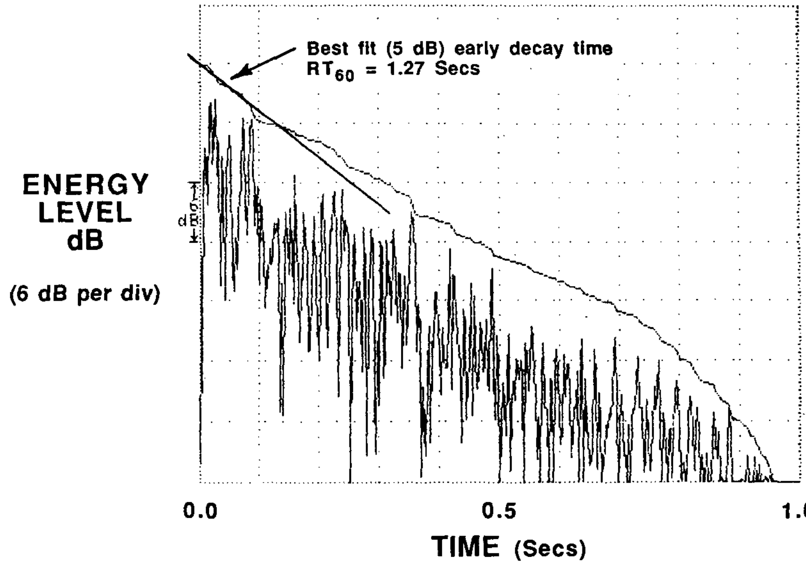


Fig. 3. The energy decay versus time (ETC) of the 2 kHz octave band for the TEF-RASTI measurement of the reverberation simulator of Fig. 2. The Schroeder reverse integration curve is also shown. The best straight line fit to the equivalent 5 dB early decay is shown. While the equivalent early decay time is 1.27 seconds, the conventionally defined RT_{60} is about 1.8 seconds.

ZERO dB SIGNAL-TO-NOISE RATIO

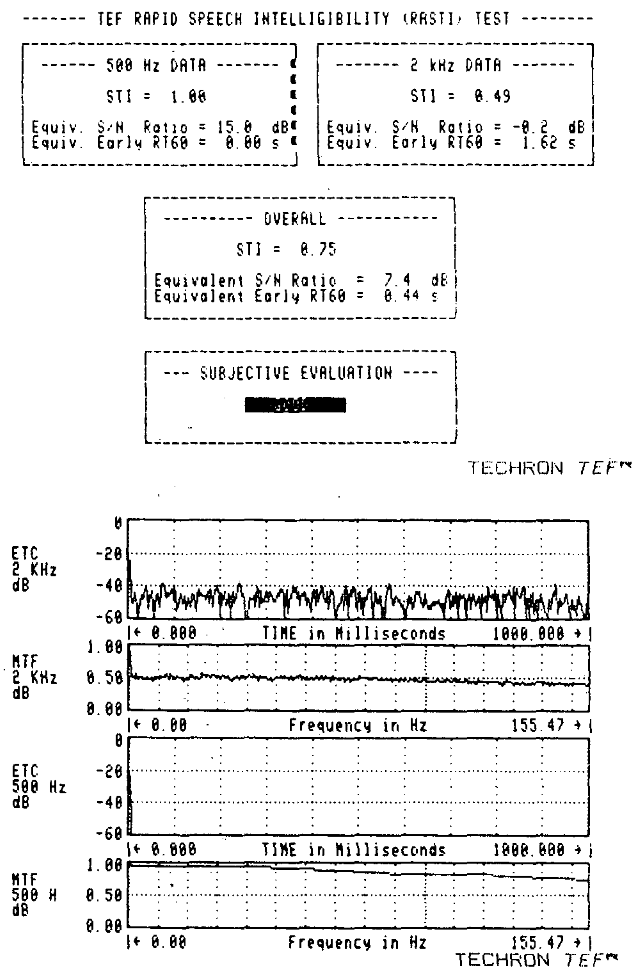


Fig. 4. TEF-RASTI intelligibility measurement on perfect transmission path but with added noise. An artificial zero dB signal-to-noise ratio condition was created by added 2 kHz one-third-octave bandwidth, pseudo-random noise to the perfect transmission path. Note that the 2 kHz TEF-RASTI STI value was reduced to roughly 0.5 by the addition of the noise. Note also that the 2 kHz MTF data was reduced by the same amount at all modulation frequencies.

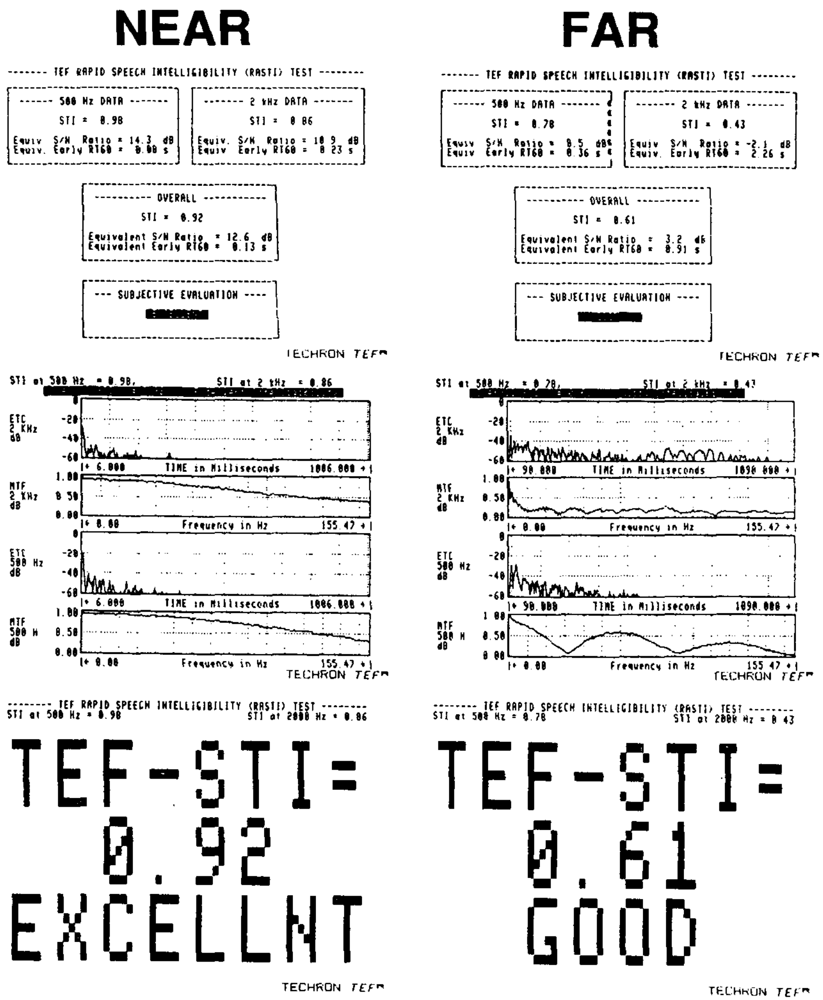


Fig. 5. Screen displays showing the results of Intelligibility measurements of a small monitor loudspeaker in a large acoustic space. Two sets of measurements were taken with the microphone to speaker distance set at 3 m (10 ft, "NEAR", left column) and 34 m (110 ft, "FAR", right column). a. TEF-RASTI data with tabular data shown at the top of the figure, graphical ETC and MTF data shown in the center, and large letter data shown at the bottom (the last "e" of excellent is missing because only 8 characters could fit across the TEF screen!). TEF-RASTI predicts an overall STI of 0.92 "Excellent" at the "NEAR" position and 0.61 "Good" at the "FAR" position.

NEAR

-------- TEF FULL STI SPEECH INTELLIGIBILITY TEST ----------

FREQUENCY	STI	EARLY RT60	S/N RATIO
125 Hz	0.74	0.45 s	7.2 dB
250 Hz	0.81	0.32 s	9.3 dB
500 Hz	0.98	0.08 s	14.3 dB
1000 Hz	1.00	0.08 s	15.0 dB
2000 Hz	0.66	0.23 s	10.9 dB
4000 Hz	0.97	0.06 s	14.1 dB
8000 Hz	0.97	0.08 s	14.1 dB

--------- OVERALL ----------

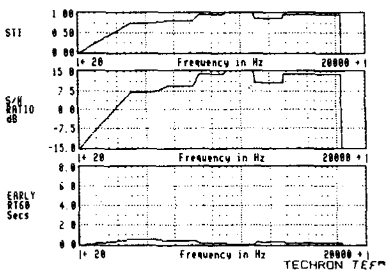
STI = 0.93

Equivalent S/N Ratio = 13.0 dB
Equivalent Early RT60 = 0.12 s

--- SUBJECTIVE EVALUATION ----

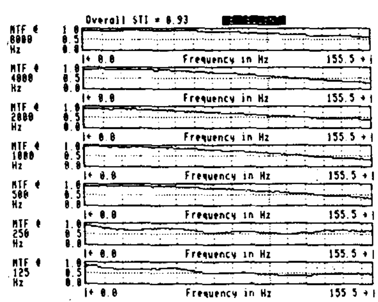
TECHRON TEF™

--------- TEF FULL STI SPEECH INTELLIGIBILITY TEST ---------
Overall S/N Ratio = 13.0 dB, Overall Equivalent RT60 = 0.12 Secs



TECHRON TEF™

Overall STI = 0.93



TECHRON TEF™

FAR

-------- TEF FULL STI SPEECH INTELLIGIBILITY TEST ----------

FREQUENCY	STI	EARLY RT60	S/N RATIO
125 Hz	0.34	3.55 s	-4.7 dB
250 Hz	0.52	1.45 s	0.5 dB
500 Hz	0.78	0.36 s	8.5 dB
1000 Hz	0.65	0.71 s	4.6 dB
2000 Hz	0.43	2.26 s	-2.1 dB
4000 Hz	0.46	1.97 s	-1.3 dB
8000 Hz	0.57	1.08 s	2.2 dB

--------- OVERALL ----------

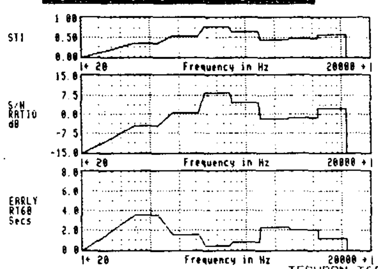
STI = 0.54

Equivalent S/N Ratio = 1.1 dB
Equivalent Early RT60 = 1.29 s

--- SUBJECTIVE EVALUATION ----

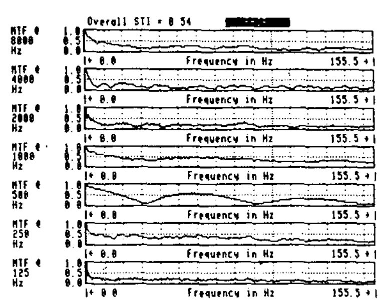
TECHRON TEF™

--------- TEF FULL STI SPEECH INTELLIGIBILITY TEST ---------
Overall S/N Ratio = 1.1 dB, Overall Equivalent RT60 = 1.29 Secs



TECHRON TEF™

Overall STI = 0.54



TECHRON TEF™

Fig. 5 Cont. b. TEF Full STI data in the large acoustic space with tabular data shown at the top of the figure; graphical STI, S/N, and early decay data versus frequency shown in the center; and graphical MTF plots at each octave center frequency at the bottom. TEF-STI predicts an overall STI of 0.93 "Excellent" at the "NEAR" position and 0.54 "Fair" at the "FAR" position. Note that the STI value at the "FAR" position for this measurement is lower than the RASTI STI value (Fig. 5a) for the same position.