Measure Speech Intelligibility with a Sound Level Meter

Ole-Herman Bjør, Norsonic AS, Lierskogen, Norway

Assessment of speech intelligibility for an emergency signalling system or a public address system may be an important task for an engineer or consultant. Today, the analysis power of a modern sound level meter is sufficient for direct measurement of speech intelligibility using STI or CIS metrics.

Several factors determine how well speech is received when we listen to a speaker. It depends on the performance of the speaker as well as on our psychoacoustical system. But, in most cases, intelligibility is characterized by the acoustics of a sound reinforcement system and associated listening space. For such applications we desire an objective assessment of the speech intelligibility, independent of the speaker and the listener. The speech transmission index, STI, has proven to be a valuable tool for such an objective assessment. From the first presentation in *Acustica* in 1971,[1] the method has been refined and diversified for various applications. Recently the International Electrotechnical Commission, IEC, has launched the third revision of the International Standard specifying the method for calculating the STI index, IEC 60268-16.[2]

The STI methods can be used to compare speech transmission quality at various listening positions and under various conditions within the same listening space. In particular it is useful for assessing the effect of changes in acoustic properties, including effects produced by the presence of an audience, changes in room surface properties or changes in a sound system. The methods are also able to predict the absolute rating of speech transmission quality with respect to intelligibility when comparing different listening spaces under similar conditions or assessing a speech communication channel.

Speech Transmission Index – STI

The STI-method for assessing speech intelligibility was developed at TNO Human Factors in the Netherlands by Tammo Houtgast and Herman Steeneken. The basis for the STI index is that the intelligibility of speech is largely based on the slow amplitude modulation of a sound that acts as a carrier. In the STI-method, the carrier is stationary gaussian noise divided into seven octave bands ranging from 125 Hz to 8 kHz. The width of each band is one-half octave. Each of the bands is modulated with one of 14 modulation frequencies. The modulation frequencies are selected in one-third octave steps from 0.63 Hz to 12.5 Hz, which gives a total of 98 combinations.

In STI context, the term "intensity" is used to denote the square of the sound pressure Pa^2. Intensity is the quantity being modulated. A small loudspeaker, roughly the size of a human mouth, driven by the modulated excitation signal, acts as a talker.

The sound at the listener position is received by a microphone. The level and degree of modulation in each octave band are used to determine the speech transmission index. Noise and reverberation in the room will reduce the observed degree of modulation. The method also considers the effect of the most common types of distortions such as harmonic distortion and intermodulation. However, some other forms of nonlinearity, like frequency shifts and frequency multiplications, are not treated effectively.

In order to take the effects of nonlinearity into account, it is important that the basic signal being modulated is a random noise signal with a high crest-factor, a spectral distribution similar to the long-term speech spectrum and that the main modulation frequencies are selected one by one. The measurement of the full STI therefore has to be performed as a sequence of measurements. If each of the 98 combinations is measured for 10 seconds, the total measurement time will be about a quarter of an hour. Such a long measurement time to make an STI measurement in only one position of a room limits the applicability of the full STI method. As a result, ways were developed to reduce the measurement time while giving almost equivalent results.

The STI method may be modified in different ways to reduce the time needed for the measurement. If the system to be measured is regarded as linear, then a number of solutions exist. The excitation signal may be modulated with all modulation frequencies simultaneously and the components may be separated after reception by the use of filters or by Fourier analysis. A more common method is to calculate the complex modulation transfer function m from the impulse response of the room from which the STI value can be determined.

If the impulse response can be regarded as a well-behaved room-response with an exponential decaying envelope, characterized by reverberation time, the modulation transfer function at frequency F may be calculated directly from the value of the reverberation time T and the effective signal-to-noise ratio S/N (in dB). A simplified formula, not taking the effects of masking and the threshold of hearing into consideration, gives the following relationship:[3]

$$m(F) = \frac{1}{\sqrt{1+\left(2\pi F \frac{T}{13.8}\right)^2}} \cdot \frac{1}{1+10^{(-S/N)/10}}$$

As seen from this formula, a limited signal-to-noise ratio reduces the modulation transfer function for all frequencies. A long reverberation time reduces the modulation mainly for the highest modulation frequencies since the formula contains the product of F and T.

The STI value is a weighted average of the different modulation indices. The last revision of IEC 60268-16 also considers masking effects and the absolute threshold of hearing.

RASTI and STIPA

In order to simplify direct measurements, the RASTI method (Rapid Analysis – or Room Acoustics – Speech Transmission Index) was developed at TNO in 1979. Different instruments were developed for the measurement according to this standard. A typical measurement time was 10 to 15 seconds. The RASTI method only considers two octave bands – 500 Hz and 2 kHz.

Due to the simplicity of use, the instruments used to determine RASTI were also used for applications beyond the main design goal – room acoustics. The RASTI value is often used for assessing the quality of public address systems, but comparisons with subjective measurements have shown that the deterioration of speech intelligibility is not handled correctly if the PA system is strongly nonlinear or suffers from limited bandwidth.

In order to improve the accuracy of the intelligibility assessment, the STIPA (Speech Transmission Index – Public Address) method was developed. It features improved handling of reverberation effects in the room and distortions commonly found in public address systems. It also works well for room acoustics and can therefore replace the RASTI method and deliver results closer to the values obtained by the full STI method in most cases.

The STIPA uses all seven octave-bands like the full STI ap-

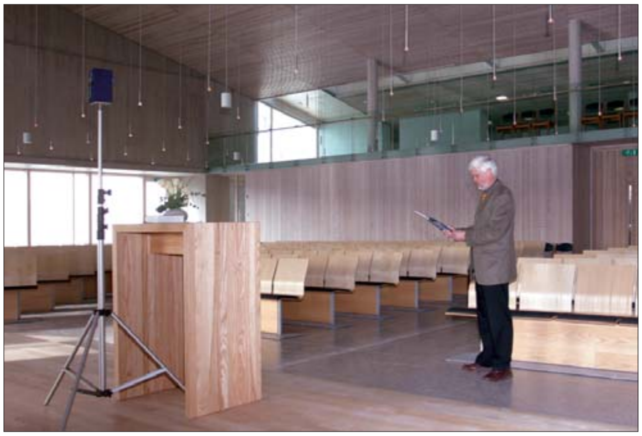

Figure 1. Example of a setup for the measurement of speech intelligibility according to the STIPA method, Vardasen Church.

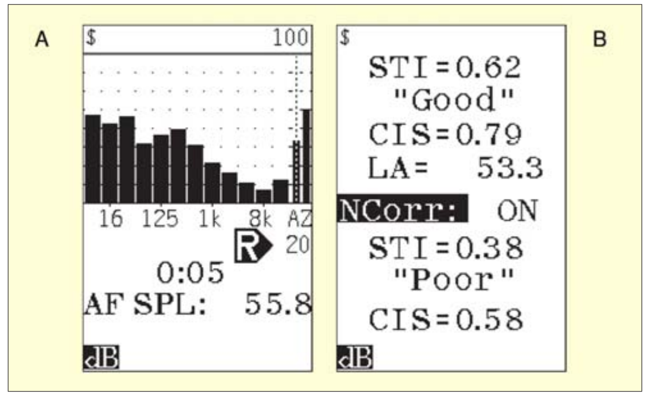

Figure 2. A – Display during measurement of STIPA. B – Screen readout after a measurement. The STIPA method is implemented as a program option in the Norsonic sound level meter N-118. Both STI- and CIS-value are presented, without and with a virtual background noise level.

proach and 12 modulation indices are measured in total. (See Table 1). The measurement time for a STIPA measurement is similar to the RASTI method – 10 to 15 sec. The STIPA method uses a higher degree of modulation for each test frequency than the RASTI method and will thus be more robust for interference from nonstationary background noise.

How STIPA is Measured

Most applications of the STIPA method require a loudspeaker to act as a talker. In this case, the directivity of the loudspeaker should be close to the directivity for a human speaker, as speech intelligibility depends upon the directivity of the source. Therefore, a mouth simulator having similar directivity characteristics to those of the human head/mouth should be used for highest accuracy when assessing the intelligibility of un-amplified talkers. Furthermore, the frequency response between 100 Hz and 10 kHz should be relatively flat. The loudspeaker should be of single element design or use a coaxial element so the acoustical center is well defined. IEC 60268-16 recommends using a loudspeaker with a maximum cone diameter of 100 mm and refers to ITU-T Recommendation P.51[6] describing an artificial mouth. A sound source according to this specification is available from different manufacturers, with a typical example being the GRAS-44AA from G.R.A.S. Sound & Vibration.

Figure 1 shows a typical set-up for a STIPA measurement in a room. The description in this article is based on the commercially available sound level meter Nor118 manufactured by Norsonic AS. STIPA is an optional mode of operation for the instrument. The instrument option comes with an audio CD containing the excitation signal – 70 minutes of male voice. The signal is the sum of the bands of noise, each modulated with two frequencies as specified for the STIPA method. The shape of the spectrum is specified in the standard (IEC 60268-16 recommends that the excitation spectrum be correct within ±1 dB

Table 1. Modulation frequencies in Hz for the seven octave bands.

Octave Band, Hz	125 & 250	500	1000	2000	4000	8000
1st Modulation freq.	1.00	0.63	2.00	1.25	0.80	2.50
2nd Modulation freq.	5.00	3.15	10.0	6.25	4.00	12.5

Table 2. Relation between the STI-value and the speech intelligibility assessment.

STI-Value	Speech Intelligibility Assessment
0.75 ≤ STI	Excellent
0.60 ≤ STI < 0.75	Good
0.45 ≤ STI < 0.60	Fair
0.30 ≤ STI < 0.45	Poor
STI < 0.30	Bad

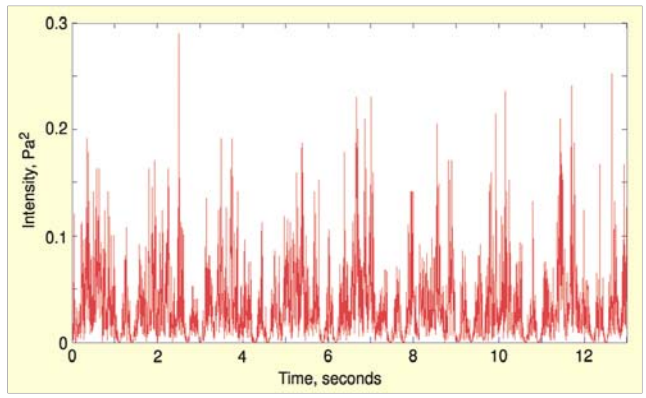

Figure 3. The sound "intensity" in the 500 Hz band computed from the sound pressure level for a measurement period of 13 seconds. The excitation is modulated with 0.63 Hz and 3.15 Hz signals.

for the applicable frequency range).

The sound level meter in the normal mode of operation may be used for verifying the level and for spectral weighting. The excitation signal runs continuously and no synchronization between the excitation and the instrument is required.

A small radio (Tivoli Audio PAL) which may be operated from internal rechargeable batteries, is used for the excitation. It comes with an input socket for the signal from a CD player, but requires the addition of a mounting bracket for use on a tripod. The diameter of the loudspeaker is about 6 cm and the frequency response is fairly flat from 100 Hz to above 10 kHz.

When the simulated speech is relayed through a sound system, an artificial mouth is normally not required unless a close talking or noise-canceling microphone is involved. If a sound reinforcement system is being used, the sound source should be placed on the axis of the appropriate microphone at a normal speaking distance (measured from the lip-circle for the artificial mouth or acoustic center of the loudspeaker) and directed in the normal speaking direction.

The sound level meter is placed in a position in the room where the speech intelligibility is to be judged. About 15 seconds are required from start to finish for the measurement. No cable is needed between the excitation and the instrument.

To test specific transmission channels, an electronic signal, instead of an acoustic signal, is needed for the input to the sound level meter for analysis. For this application, the microphone preamplifier is replaced by a cable to the input socket of the sound level meter.

The following description applies to the Norsonic Nor118: To make a measurement, switch on the sound level meter and ensure that it is properly calibrated. Play the STIPA excitation signal and adjust the level to the required speech level, normally about 60 dB A-weighted, at 1 m from the sound source (66 dB at 0.5 m). The excitation is played continuously through the loudspeaker. The instrument is ready for measurement because all parameters for controlling the measurement are preset and the user only has to press the start button. Figure 2A shows the display on a meter during measurement and Figure 2B shows the end results. The STI value, calculated according to the STIPA method, is shown on the screen together with the assessment: 'Excellent,' 'Good,' 'Fair,' 'Poor' or 'Bad' (see Table 2). Also, a tabular display showing the mean level in each band can be displayed as well as each of the 12 measured modulation indices. The corresponding CIS-value (Common Intelligibility Scale) according to IEC 60849 (1998-02) "Sound Systems for Emergency Purposes" is also indicated.[5] There is a nonlinear relation between the STI and CIS value although both use 1 to indicate the best intelligibility and 0 for the poorest.

During the measurement, the short time equivalent level in each octave band is measured with a time resolution of 5 msec. The level is converted to squared sound pressure, Pa^2, or 'intensity' in this context (see Figure 3). The degree of modulation for each modulation frequency is found by Fourier transformation. By comparing the measured degree of modulation with the degree of modulation in the excitation (standard at 55%), the value for the modulation transfer function is obtained. In total, 12 combinations of carrier/modulation frequencies are measured. The measured modulation transfer function is corrected in order to compensate for the threshold of hearing and for the masking effects of the human auditory system.

In practice, the CD player, loudspeaker and sound level meter combination needs to be checked under nearly ideal conditions. Such a situation exists just in front of the loudspeaker placed in a large room when the background noise is considerably lower than the excitation signal. If the CD player has a shock protection system it is normally recommended to switch it off as it often reduces the obtainable STI value. If the level is adjusted to fall somewhere between 60 and 70 dB, the STI value should be close to 1.

Analysis and Interpretation of the Results

It is important to examine the modulation matrix to determine the reliability of the results. As a rule, the modulation index in each octave-band should decrease with increasing modulation frequency. Constant or slightly decreasing values in a column indicate the presence of noise. Large reductions indicate that reverberation is the main effect. Values that increase with increasing modulation frequency indicate the presence of periodic or strong reflections, which may produce an over-optimistic conclusion. If this effect is detected, it should be reported with the results and an estimated correction applied.

Background and Occupancy Noise

The STIPA method considers the effect of the actual background noise when intelligibility is assessed. If a real background noise is present during the measurement, the effect of the noise will in most cases be treated as noise according to the STIPA method and lower the STI value.

Some nonstationary noise signals may be interpreted as a STIPA excitation signal, wrongly giving a modest STI value. The instrument is programmed to detect such situations and place a question mark adjacent to the indicated value if this occurs. However, not all cases can be detected properly. To investigate such potential problems, we recommend making a measurement without the excitation signal and only the background noise. If this STI value is low or considerably lower than the value measured with the excitation signal (preferably STI < 0.2), the measurement will have a high degree of reliability.

If the background noise can be switched off, it is possible to measure the response to the STIPA excitation and noise level separately. This will eliminate the interference between a spiky-noise signal and the STIPA method.

Sometimes we want to find out what the intelligibility would be with a certain background noise. A typical application is to assess the intelligibility of an auditorium with an audience to the auditorium as measured without an audience. For such applications, the noise may be accounted for by either using a previously measured spectrum or by keying in an artificial expected spectrum.

The correction is done according to the following equation:

$$mc_{k,f} = m_{k,f} \frac{I_k}{I_k + Irs_k + Ino_k + Iam_k}$$

where

$mc_{k,f}$ = the corrected modulation transfer function for octave band number k and modulation frequency f.

$m_{k,f}$ = the measured modulation transfer function (as displayed).

I_k = the measured 'intensity' in octave band number k.

Irs_k = the 'intensity' in octave band number k related to the threshold of hearing.

Ino_k = the optional 'intensity' in octave band number k corresponding to a specified background noise level. If not used this value is zero.

Iam_k = an 'intensity' in octave band number k used to mimic the masking effect in the human auditory system. The value is a function of the level in the adjacent lower octave band.

Note that the noise correction is an extension of the method specified in IEC 60268-16. This allows a measurement in a situation without background noise and recalculation of what the STI value would have been with a specified noise level. The correction may be done directly in the instrument, or the results may be downloaded to a PC and post-processed with the accompanying Excel® spreadsheet STIPA-Calc (as provided by Norsonic).

Precision of the STIPA Method

Because the test signal is band-limited random or pseudorandom noise, repetition of measurements does not normally produce identical results, even under conditions of steady interference. The results center on a mean with a certain standard deviation. Typically, the standard deviation of the STIPA value is about 0.02 with stationary noise interference. With fluctuating noise (for example, a babble of voices), higher standard deviations may be found, possibly with a systematic error. Repeating measurements for at least a restricted set of conditions should provide an estimate of the standard deviation.

Limitations of the STIPA Method

The STIPA method should not be used for public address systems that:

- Introduce frequency shifts or frequency multiplication.
- Include vocoders (i.e. LPC, CELP, RELP, etc.). [A vocoder is a device (voice-coder) for recreating a speech signal from information from the slow modulation of the envelope of a speech-signal. This procedure is used by the military to reduce the needed bandwidth for telephone conversations.]
- Have a background noise that is impulsive.
- Introduce strong nonlinear distortion components.

Application Example

The STIPA method was used to assess the speech intelligibility in a recently inaugurated church, Vardasen, in the vicinity of Oslo. The sanctuary of the church has an area of approximately 440 m^2 with 285 seats. The PA system is based on six distributed loudspeakers placed about 4 m above the floor. The reverberation time for the room, when empty, varies from about 1.5 sec at 125 Hz, to a maximum of 2.0 sec at 1 kHz and to a minimum of 1.0 sec at 8 kHz. From the reverberation time alone, an STI value of about 0.48, corresponding to fair intelligibility, should be expected, assuming that the level is sufficiently above the background noise.

A STIPA measurement was performed with the handheld sound level meter Norsonic Nor118 equipped with the program

Figure 4. Norsonic N-118 real time analyzer.

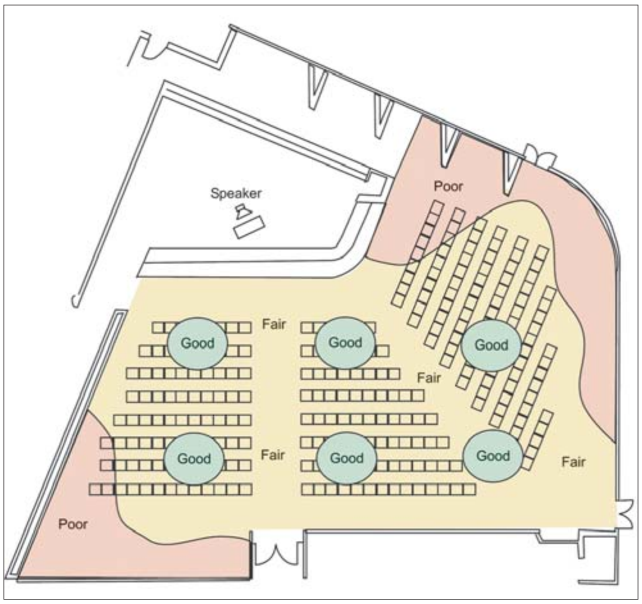

Figure 5. The speech intelligibility in Vardasen Church. The intelligibility was found to be good just below the loudspeakers and fair in most other places in the room.

option for such measurements (see Figures 1 and 4). The excitation was provided by a small commercially available radio (Tivoli PAL) playing the signal recorded on the CD Nor1034 accompanying the instrument option. The radio has a single element loudspeaker with a diameter of approximately 6 cm, and has a fairly flat frequency response from 100 Hz to well above 10 kHz. The radio was placed in the normal position for a human speaker.

All of the test equipment is battery operated. The excitation signal runs continuously and the measurements are done consecutively for the various locations in the room. The measurement time for each position is 13 seconds and the result is stored automatically after each measurement. The result consists of the STI value, the A-weighted speech level, the octave-band levels, and the modulation matrix (the modulation indices for each of the twelve modulation frequencies). The values may also be corrected for an assumed background noise level

The results are shown graphically in Figure 5. Most of the room measurement results were assessed as 'Fair' intelligibility corresponding to an STI value between 0.45 and 0.6. Two sections of the seating area, farthest away from the loudspeaker were assessed as 'Poor.' The lowest value was 0.42. The highest STI values were obtained just below the loudspeakers. The maximum STI of 0.66 corresponded to an assessment of 'Good' intelligibility. The measurements show that the main effect of the PA system is the enhancement of speech level. The increase in intelligibility from less reverberant sound is only achieved just below the loudspeakers.

Concluding Remarks

Implementation of the STIPA measurement mode in an ordinary sound level meter allows an easy and quick measurement of speech intelligibility. Normally, the STI-value is measured with normal background noise and the value indicates the speech intelligibility for this situation. For situations where the measurement is performed in a nontypical quiet environment, a synthetic background noise may be added. The STI value will then take this situation into account. This feature is especially useful when assessing speech intelligibility in auditoriums measured without an audience.

References

1. Houtgast, T. and Steeneken, H. J. M., "Evaluation of Speech Transmission Channels by Using Artificial Signals." *Acustica*, vol. 25, pp 355-367, 1971.
2. IEC 60268-16 (2003-05), "Sound System Equipment – Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index."
3. Houtgast, T., Steeneken, H. J. M. and Plomp, R. "Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. I. General Room Acoustics," *Acoustica*, 46, 60-72, 1980.
4. Bjor, O. H., "STIPA – The Golden Mean Between Full STI and RASTI," CFA/DAGA '04, Strasbourg, 2004.
5. IEC 60849, "Sound Systems for Emergency Purposes." Ed. 2.0 1998-02.
6. ITU-T Recommendation P.51: 1996, Artificial Mouth.

The author can be contacted at: ohbjor@norsonic.com.