BNAM 2021
Baltic-Nordic Acoustics Meeting
3rd-5th of May, Oslo, Norway

# A review of STI measurements

Henrik Möller
Akukon Ltd, Helsinki, Finland henrik.moller@akukon.com

Since the STI concept was introduced by Houtgaard and Steeneken in 1973 [1], the different derivations of the STI has been the principle parameter for evaluation of speech intelligibility, both in sound reinforcement systems as well as in room acoustics.

In current software implementations, at least 5 different variations of the basic STI parameter can be found. STI is described as a requirement (as a measured parameter) in several international standards. However there seems to be some differences in how the parameter is measured and calculated.

In this paper, a review of some of the measurement techniques and calculation methods will be presented and a variation of the STI parameter, based only on omnidirectional sources and impulse response measurements will be presented.

## 1 Introduction

Simplified, speech intelligibility is dependent on two acoustic factors: Background noise level and reverberation. Both these factors will decrease the signal to noise ratio.

Traditionally, speech inteligibilty of different communications channels was measured directly, that is by reading words or speech like sound, and listeners would write what they heard. Obvious this procedure was rather time consuming, so a faster, objective methods was needed. One of the first measurement methods was % Articulation, defined by V.O. Knudsen and calculated from the reverberation time, the speech level and the Signal-to-Noise ratio:

$$\%\text{Articulation} = 96 * K_t K_l K_n \quad (1)$$

Where: $K_t$ is derived from the revereberation time

$K_l$ is derived from the SPL (speech)

$K_n$ is derived from the background noise

Articulation Loss of Consonants, $\%AL_{cons}$ was introduced by V.M.A. Peuts in 1971. Percentage articulation loss of consonants can be calculated by:

$$\%AL_{cons} = \frac{200 D_2^2 RT_{60}^2}{VQ}. \quad (2)$$

The advantage of the $\%AL_{cons}$, is that it can be be estimated from statistic acoustic parameters and relatively easily be measured from the direct sound level, the reverberant sound level and the background noise level.

Davis and Davis [4] established an experimental connection between RASTI(STI) and %ALcons:

$$\%AL_{cons}=170{,}5405e^{-5{,}419RASTI} \quad (3)$$

$$RASTI=0{,}1845LN(\%AL_{cons})+0{,}9482 \quad (4)$$

# 2 Speech intiligibility parameters

The STI and related parameters are defined in IEC60268-16:2011 [4]. In the standard, the parameter is defined for sound reinforcement system, however not described as a room acoustic parameter. In the standard ISO 3382-3:2012, the inverse STI is used as measure of speech privacy. The standard requires that the measurement must be done with an omnidirectional loudspeaker.

## 2.1 STI

The Speech Transmission Index (STI) is defined from the Modulation Transfer Function. This is calculated for the intensity envelope of a signal.

The value is calculated from the octave bands 125 Hz to 4000 Hz and for 14 modulation frequencies from 0,63 Hz to 12,5 Hz (third octave bands).

$$m(F) = \frac{1}{\sqrt{1+\left(2\pi F\frac{T}{13.8}\right)}} \cdot \frac{1}{1+10^{(-S/N)/10}} \quad (5)$$

In general, both the modulation transf er function and the STI can be calculated for the intensity envelope of any signal.

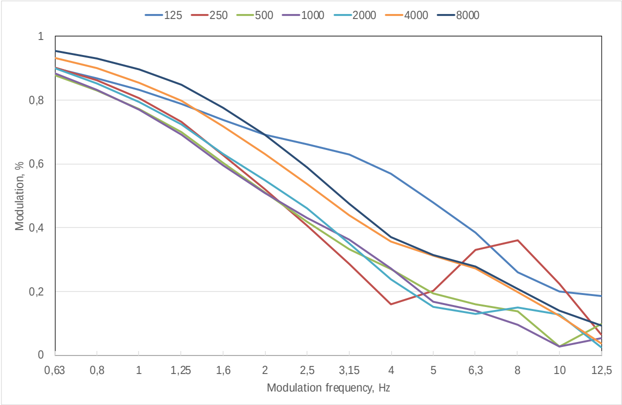


Figure 1: Example of MTF curves derived from an impulse response measurement. The STI-value is 0,48

## 2.2 RASTI

However, as a full STI measurement and calculation was quite calculation heavy, therefor a simplified parameter, the RApid Speech Trasmission Index was proposed. For these 4 modulation frequencies is calculated for the 500 Hz band and 5 modulation frequencies for the 2000 Hz band.

The first practical measurement device was the Brüel&Kjaer Type 3361 Rasti meter. This used a modulated pink noise signal (500 Hz and 2000 Hz octaves) . The modulation frequencies were 1, 2, 4 and 8 Hz for the 500 Hz octave and 0.7, 1.4, 2.8, 5.6 and 11.2 Hz for the 2000 Hz octave. In the latest version of the IEC 60268-16, RASTI has been declared obsolete, and should no longer be used.

## 2.3 STIPA

As measurement equipment (computers) became more powerful, the STI measurement was integrated into most measurement software, where typically the STI would be calculated from the impulse response. However, measuring the impulse response from a PA system with time variance, non-linearities etc is far from straight forward, hence a system to check speech intelligibility for PA-system was needed. The STIPA was described for this purpose.

Instead of the 14 modulation frequencies applied to all seven octave bands, the STIPA method two unique modulation frequencies simultaneously to each of the seven bands. Hence the STIPA uses 2x7=14 modulation frequencies. Also, it is stated that the STIPA method is only valid for male speech spectrum.

## 2.4 STITEL

In IEC60268-16 (2011) also a simplified STITEL method is described, but as far as this author knows, it is not very widely used.

# 3 BASIC ATTRIBUTES OF HUMAN SPEECH

From a speech intelligibility point of view, the main attributs of human speech are the frequency spectrum and sound level of speech as well as the directivity of the head/mouth. It should also be noted that in most languages, the consonants are essential to speech intelligibility. This means that the higher frequencies, in particular the 2 000 Hz octave band, holds the greatest amount of information, see fig 1.

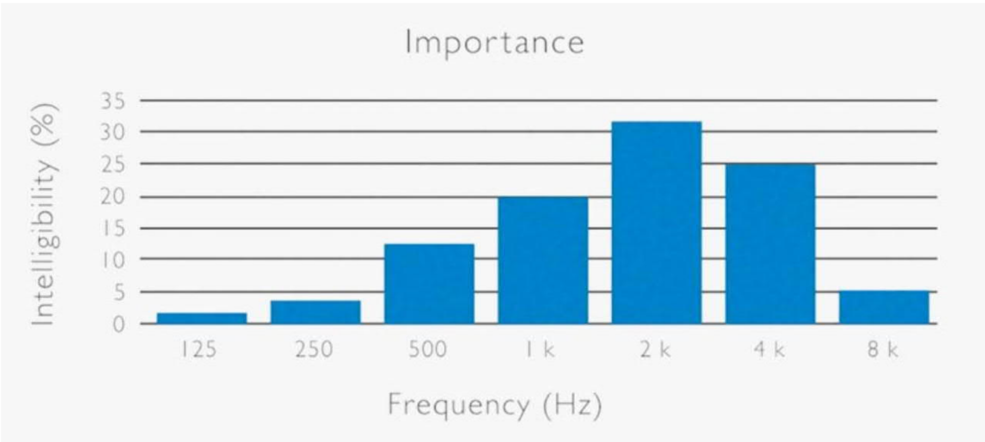


Figure 2. Importance to intelligibility for 125 – 8000 Hz octave bands [3]

However, as can be seen from Figure 2, the human speech frequency spectrum is not very "high frequency", essentially it is only when we are shouting, that there is a shift towards more high frequency energy at high frequencies.

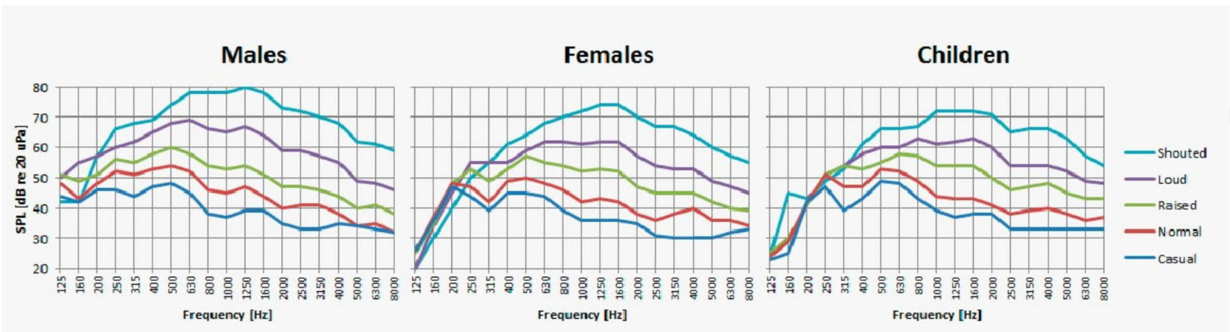


Figure 3. Voice spectra (1/3 octave) depending on efforts [3]

The directivity of human speech, averaged for both sexes, is shown in Figure 4. It is obvious that mainly higher frequencies show some directivity, whereas below 500 Hz, the human voice is more or less omnidirectional.

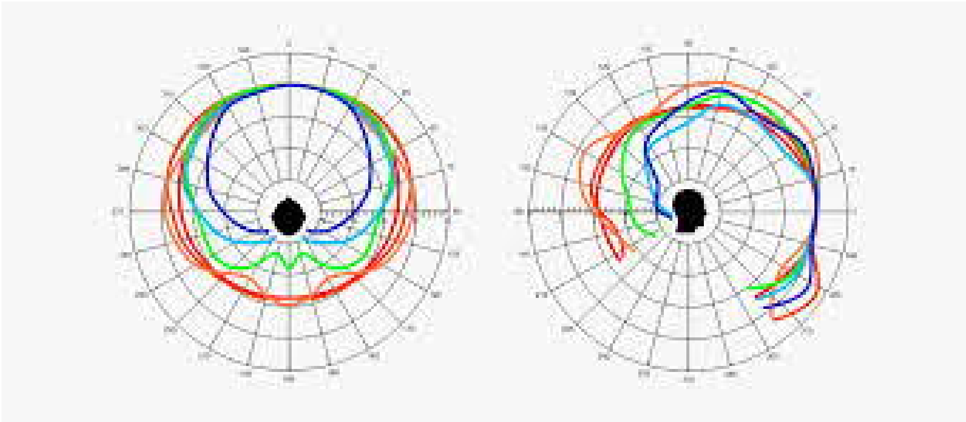

Figure 4. Human talker, polar plots octave intervals, from 250 – 8000 Hz. Division 5 dB. [2] [3]

# 4 Standardized use

## 4.1 IEC 60849:1998

In standard for sound systems for emergency purposes, the requirement is that intelligibility should be at least 0,7 on the common intelligibility scale (CIS). In appendix there are conversion graphs between CIS and other intelligibility measures.

## 4.2 ISO 3382-3:2012

In the standard for acoustic quality for open plan offices, STI is used as a measure of privacy and distraction. The standard states that the STI should be measured with a sound level calibrated omnidirectional source. The Distraction distance is defined as the distance where STI falls below 0,5 and the Privacy distance is the distance where the STI falls below 0.2.

## 4.3 Other use

Some countries also use STI parameters as a requirements or suggestion, for classrooms in schools and other spaces for education. For instance, in the Finnish standard SFS5907 Acoustic classification of spaces in buildings, it is defined that the minimum requirement for STI in classroom is ≥ 0,7 and the ≥ 0,8 is what should be aimed for. The requirement for classrooms for children with poor hearing or speech impediments is ≥ 0,75 and the goal to aim for is ≥ 0,85.

Also, in the requirements for sport events (FIFA, UEFA etc) there are requirements for speech intelligibility is stadiums. These requirements are typically in line with the IEC 60849 requirements.

# 5 Measurement techniques

There are defined two measurement methods for STI related parameters:

## 5.1 Direct STI method

This method uses a modulated (speech like) test signals to directly measure the modulation transfer function. Typically modified Pink Noise with modulation frequencies. In this case, the measurement signal is either applied as an electric input to the system or through a “human speaker” loudspeaker to a microphone.

The response is picked up via an omnidirectional microphone.

## 5.2 Indirect STI method

This method uses the impulse response and forward energy integral (Schroeder integral) to derive the modulation transfer function. This that STI related parameters can be measured at the same time as other room acoustic parameters. This means that the speech intelligibility will normally be measured using an omni-directional speaker and the response measured with an omnidirectional microphone.

## 5.3 Equipment requirements

In the standard IEC60268-16 (2011) it is stated that the microphone should be an omnidirectional, diffuse field microphone. Also, the loudspeaker should ideally be an artificial mouth, but also a small loudspeaker, element less than 100 mm, can be used. In both cases the sound source should be calibrated, both with respect to level and frequency to have the characteristics of human voice.

As one of the factors determining speech intelligibility is background noise level, or more precisely the signal to noise level, the directivity of the loudspeaker used for the measurement has a large influence. Basically, by increasing the Q-factor for 1 (omnidirectional) to 2, the soundlevel on axis rises by 3 dB.

$$\Delta L = 10LOG\left(\left(\frac{Q}{4\pi r^2}\right) + \frac{4}{A}\right) \quad (6)$$

This would mean that the STI measured using omnidirectional speakers, would have somewhat low correlation to actual speech intelligibility for higher frequencies.

# 6 Simulations of the STI

Typically, STI will be simulated using a room acoustic simulation program, such as Odeon or Catt. For a real STI simulation, a human talker source is used, and the background noise level is adjusted to appropriate levels.
In the latest version of the IEC60268-16 (2011) there is however also described a procedure to estimate STI from statistical acoustic parameters. The input is the reverberation time and background noise level, all data in octave bands.

$$m(f_m) = \frac{\sqrt{A^2+B^2}}{C} \quad (7)$$

$$A = \frac{Q}{r^2} + \frac{1}{r_c^2}\left[1 + \left(\frac{2\pi f_m T}{13,8}\right)^2\right]^{-1}$$

$$B = \frac{2\pi f_m T}{13,8 r_c^2}\left[1 + \left(\frac{2\pi f_m T}{13,8}\right)^2\right]^{-1}$$

$$C = \frac{Q}{r^2} + \frac{1}{r_c^2} + Q10^{-SNR/10}$$

Where: $f_m$ is the modulation frequency
r is the source-receiver distance
$r_c$ is the critical distance
T is reverberation time.
SNR is the signal to noise level, ie the difference between normalized speech and the measured or predicted background noise level

# 7 Discussion

Originally STI related parameters was intended for PA systems, that is the measurement signal was fed to the PA-system and STI was measured in different places in the space.

However, most likely most measurements of STI related parameters are currently done as a part of general room acoustics measurement, that means using an omnidirectional loudspeaker and calculate the STI according to the indirect method.

So, it seems that there are a large variety of both STI parameters as well as measurement methods. But again, the most common way to measure STI, as part of the measurement of room acoustic parameters using impulse responses is not clearly defined.

The main problem is that as STI is dependent on the signal-to-noise ratio, and therefore should be measured with a sound source calibrated to 68 dB (speech level), the sound level used for room acoustic measurements are some louder, in order to ensure sufficient signa-to-noise ratio for the room acoustic parameters.

## 7.1 Suggestion for geometric STI

To accept these facts, a new STI index is proposed, $STI_{geo}$. This is essentially the STI measured as part of a standard room acoustic parameter measurement.

This should be measured with an omnidirectional microphone and an omnidirectional loudspeaker, as defined in ISO 3382-1. However, the calibration of the loudspeaker should be in accordance to the requirements of the Strength parameter, not STI. The measurement will typically be done using swept sine signal.

This will perhaps be higher than "normal" STI as background noise level is not considered, and lower at high frequencies where the human voice is more directive than the omnidirectional loudspeaker.

However, it will still give a good indication of the speech intelligibility from a pure room acoustic point of view. This would be an easily comparable parameter, as it removes the influence of a non – ideal "human speech" loudspeaker as well as the influence of background noise.

# 8 Conclusion

In this paper, a review of the different standards and definition of STI and related parameters are presented. It is clear that there are quite a variety of measurement methods, which will in most circumstances, will give variating results.
It is therefore suggested to use the parameter $STI_{geo}$, defined as the STI measured with a measurement setup in accordance with ISO 3382-1. This will not, of course replace the STI measured for specific requirements, such as emergency systems, but could replace for instance the STI requirement is for classroom acoustics and similar.

# 9 REFERENCES


[1] T. Houtgast and H.J.M. Steeneken: "The Modulation Transfer Function in Room Acoustics and a predictor of Speech Intelligibility", Acustica 28 PP 66-73, 1973.

[2] W.T. Chu and A.A.C. Warnock: Detailed Directivity of Sound Fields Around Human Talkers, NRC-CNRC IRC-RR-104, 2002 (https://nrc-publications.canada.ca/eng/view/fulltext/?id=c1449aba-ee5b-48de-8312-ec325d31ef37)

[3] DPA Mic University, Facts about speech intelligibility, https://www.dpamicrophones.com/mic-university/facts-about-speech-intelligibility

[4] C. Davis and D. Davis: Sound System Engineering, Second Edition, Focal Press, 1982.