APPLICATION NOTE

Measuring Speech Intelligibility using DIRAC Type 7841

Speech intelligibility is an important measure of the effectiveness or adequacy of a communication system or to communicate in a noisy environment. In many daily life situations it is important to understand what is being said, for example over a loudspeaker system, and to be able to react to acoustic signals of different kinds.

Amplified speech systems like public address systems, telephones, radio links, intercoms are vital to society, but even unamplified unaided speech is important in offices, workshops, vehicles and many other situations.

With Brüel & Kjær's state-of-art equipment, speech measurements can be carried out through an artificial mouth-directional loudspeaker sound source or through direct injection into a sound system while still taking into account the impact of external and internal noise sources. Objective evaluations using one or more standardized methods can then be performed and the causes for lowered speech intelligibility identified.

DIRAC

DIRAC PC software is used for measuring a wide range of room acoustical parameters. Speech intelligibility measurements can be carried out in compliance with the IEC 60268–16 standard through an artificial mouth-directional loudspeaker sound source or through direct injection into a sound system, taking into account the impact of background noise. For accurate measurements according to the ISO 3382 standard, internally or externally generated stimulus signals through a loudspeaker sound source can be used.

Incorporating Brüel & Kjær room acoustic solutions, it is possible, among other things, to:
- Determine whether speech intelligibility in, for example, a church or railway station is sufficient
- Determine how many loudspeakers are needed to raise the speech intelligibility to a satisfactory level
- Determine whether speech intelligibility is dictated by reverberance or background noise
- Determine how well performers on stage can hear themselves and others
- Refine and troubleshoot a structure's acoustical design parameters to improve speech intelligibility

DIRAC is a valuable tool not only for field and laboratory acoustics engineers, but also for researchers and educational institutions. To learn more about DIRAC and its applications, please visit www.bksv.com/DIRAC.

Speech Intelligibility Basics

Speech transmitted across a room by a person or a public address system is never received at a listening position as an exact replica of the original signal. Speech intelligibility in everyday conditions depends on the:
- level of background noise
- distance from the speaker to listener
- loudness of the speech (signal strength)
- voice spectrum of the speech
- amount of reverberation (echoes)

Brüel & Kjær

The intelligibility is adversely affected by noise. To achieve full sentence intelligibility for listeners with normal hearing, the signal to noise ratio – the difference between the speech level and the sound level of the interfering noise – should be at least 15 dB(A). The nature of speech sounds determines the mechanism of loss of speech intelligibility. Vowels and consonants convey different sound energy. Consonants are spoken more softly than vowels, and tend to get drowned out in noisy environments. The average level of consonants is 10 – 12 dB lower than the level of vowels.

Table 1
Average vocal effort and sound level

Vocal Effort	dB(A)
Whispering	32
Soft	37
Relaxed	42
Normal (private)	47
Normal (public)	52
Raised	57
Loud	62
Very Loud	67
Shouting	72
Max. Shout	77

Consonants are more important than vowels in understanding speech. They convey most of the word information, acting as breakpoints, separating syllables and words from one another. Moreover, the most energy of consonants is at higher frequencies, while vowels are represented at lower and middle frequency bands.

The sound signal of speech is rich in temporal and frequency patterns. The speech band covers the range from around 125 Hz – 8 kHz, with the most important cue bearing energy being between 300 and 3000 Hz, and the waveform envelope varies in amplitude. These fluctuations of power in time and frequency are called modulations. Speech intelligibility is to a large extent based on this slow modulation of a sound pressure signal and despite their acoustic complexity, spoken words remain intelligible after drastic degradations in either time or frequency.

Objective Methods and Parameters

Speech Intelligibility is usually expressed as a percentage of words, sentences or phonemes (speech sounds making up words) correctly identified by a listener or group of listeners when spoken by a speaker or a number of speakers. However, such subjective intelligibility tests are time consuming and expensive and it is possible to estimate speech intelligibility from physical measurements and dispense with the need for live test subjects. The standards of the International Electrotechnical Commission (IEC) and the International Standards Organization (ISO) incorporate such objective methods for evaluating speech intelligibility.

Physical measurement techniques make use of a synthesised speech-like waveform played through an artificial voice device or small loudspeaker to act as the source for the tests. This can also be played through a PA system or sound-field system. A microphone placed at the listener position receives this signal and from the degree of modulation in each band the speech transmission is calculated. Noise and reverberation in the room will reduce the observed degree of modulation and affect the result.

STI, RASTI or STIPA are the most established parameters for measuring speech intelligibility. All of them basically apply the same principle, whereby RASTI and STIPA are a simplified version of STI. They are all based on measuring the MTFs (Modulation Transfer Functions) in seven octave bands. For each octave band there is one MTF quantifying the preservation degree of the intensity modulations in this band. These functions quantify how much the intensity modulations are preserved in seven octave bands covering the long-term speech spectrum.

A typical speech intelligibility measurement can be carried out as follows:

1. A stimulus is played through an artificial mouth-directional sound source at the talker position. E-sweeps are more robust, while MLS signals sound less obtrusive.
2. The responses are recorded in DIRAC, either directly for immediate results or through a hand-held sound recorder for higher portability.
3. Back at the office it is possible to see the effect on speech intelligibility if the signal and/or background noise levels are changed.

Modulation Transfer Function (MTF)

The speech intelligibility parameters in DIRAC are based on the relation between perceived speech intelligibility and the intensity modulations in the talker's voice, as described by Houtgast et al. [1, 6]. When a sound source in a room is producing noise that is intensity modulated by a low frequency sinusoidal modulation, the modulation depth at the receiver position will be reduced due to room reflections and background noise. The modulation transfer function (MTF) describes to what extent the modulation is transferred from source to receiver, as a function of the modulation frequency F, which ranges from 0.63 to 12.5 Hz. Hence, the MTF depends on the system properties and the background noise (see Fig. 1).

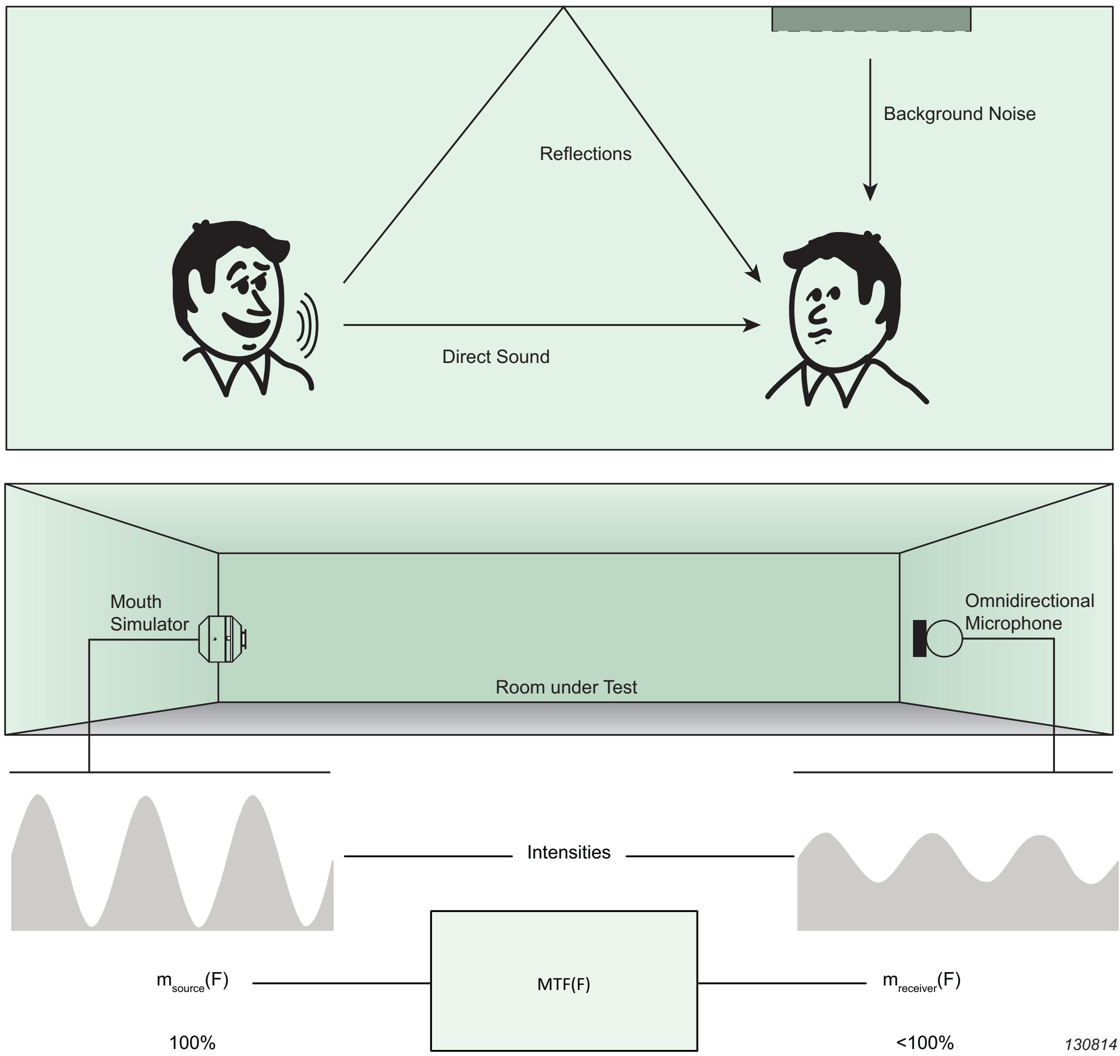


Fig. 1
Relation between speech intelligibility and modulation depth

Modulation Reduction Matrix

To evaluate speech intelligibility the MTF is determined for each octave frequency band relevant for speech. High MTF values indicate a good transfer of the level modulations and hence a good speech intelligibility. Low MTF values indicate a significant reduction of the speech intelligibility, due to the acoustical system properties and/or background noise.

The MTF values for the 14 modulation frequencies are averaged, resulting in the modulation transmission index (MTI). The modulation transmission indices for the 7 octave band frequencies can be processed to arrive at the Speech Transmission Index STI (see also [1]). Table 2 shows an example of a modulation reduction matrix.

From the modulation reduction matrix, you can obtain information on the cause of the reduction of the speech intelligibility. A constant MTF over F indicates background noise, a continuously decreasing MTF indicates reverberation and an MTF first decreasing and then increasing with F indicates an echo. For example, the modulation reduction matrix of Table 2 reflects a case where only reverberation plays a role, while the influence of background noise and echoes is negligible.

Table 2
Example of STI modulation reduction matrix

Modulation Frequency F [Hz]	Octave Band Frequency [Hz]						
	125	250	500	1k	2k	4k	8k
0.63	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.80	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.60	0.858	0.866	0.806	0.852	0.837	0.841	0.835
2.00	0.858	0.866	0.806	0.852	0.837	0.841	0.835
2.50	0.858	0.866	0.806	0.852	0.837	0.841	0.835
3.15	0.858	0.870	0.806	0.852	0.837	0.841	0.835
4.00	0.858	0.866	0.806	0.852	0.837	0.841	0.835
5.00	0.651	0.676	0.543	0.664	0.630	0.633	0.612
6.30	0.651	0.676	0.543	0.664	0.630	0.633	0.612
8.00	0.506	0.533	0.377	0.531	0.506	0.502	0.471
10.00	0.506	0.533	0.377	0.531	0.506	0.502	0.471
12.50	0.444	0.460	0.267	0.427	0.406	0.413	0.373
MTI	0.75	0.76	0.70	0.75	0.73	0.74	0.73

The used octave frequency bands are related to the typical frequency range of a human voice. A differentiation between male and female voice spectra is made in IEC 60268-16 [2]. The female voice spectrum model does not include the 125 Hz octave band.

Measuring MTF: Modulated Noise Versus Impulse Response

Two commonly used MTF measuring methods are the modulated noise method and the impulse response method.

In the modulated noise method the excitation signal basically consists of 7 × 14 = 98 summed noise signals, each of which is filtered and modulated according to the matrix in Table 2. The signal is picked up at the listener position and for each octave band and modulation frequency (F) the modulation reduction MTF (F) is measured. A complication for the full STI method described here is that the modulations in one octave frequency band can influence the modulations in other frequency bands. Not all 98 modulations can therefore be measured at once. Also, due to the randomness of the excitation signal, it takes a relatively long time to obtain reproducible results. In practice a single full STI measurement requires at least 15 minutes. The receiver can also misinterpret background noise fluctuations as signal modulations, and overestimate the speech intelligibility at low SNR values.

Schroeder [3] has shown that the MTF, can also be derived from the Fourier transform of the squared impulse response. Rife [4] has used [1] and [3] to include the impact of background noise. If the impulse response is measured through deconvolution of a deterministic signal, such as an MLS or sweep signal, the measurement takes far less time than with the modulated noise method for the same reproducibility (some 5 s on average). The impulse response method does, however, require more processing power.

MTF Measurement Conditions and Limitations

The MTF as a basis for the speech intelligibility also has its limitations. Distortions in the system under test may affect the MTF (hence the measured speech intelligibility) differently from the real speech intelligibility. For instance, a recorded voice that is played back at a slightly higher speed is still very intelligible, but the measured MTF may drop significantly. Centre clipping (cross-over distortion) may affect the real speech intelligibility much more severely than the measured MTF. The same holds for signal drop outs (Houtgast, Steeneken et al. [6]). In general the deviation between real and measured speech intelligibility will be different for the two measuring methods mentioned in the previous section.

In the IEC 60268-16 standard [2], some conditions are given to avoid problems:

1. The system under test should not introduce frequency shifts or use frequency multiplication.
2. The system under test should not contain vocoders, such as LPC, CELP and RELP.
3. The speech transmission should be essentially linear, with amplitude compression or expansion limited to 1 dB, and no peak clipping.

Obviously, we can add:

4. The system under test should not introduce centre clipping.
5. The system under test should not introduce drop outs.

It is therefore important to be aware of any nonlinear behaviour when measuring the speech intelligibility through a sound system. If the system is behaving linear, the measured and real speech intelligibilities correlate very well for both methods.

Parameters Related to Speech Intelligibility

Speech Transmission Index (STI)

The speech transmission index STI is the most comprehensive and important speech intelligibility parameter in DIRAC. Although not usable for transmission channels that introduce frequency shifts or frequency multiplication, or include vocoders (voice encoders), the STI takes into account most effects that could cause deterioration of the speech intelligibility. For details, refer to [2].

Technically, the STI is calculated as the weighted sum of modulation transfer indices MTI, one for each octave frequency band from 125 Hz through 8 kHz (where each MTI value is derived from MTF values over 14 different modulation frequencies, see Table 2) taking into account auditory effects according to IEC 60268-16.

Table 3
Relation between STI and speech intelligibility

STI	**0.00 – 0.30**	**0.30 – 0.45**	**0.45 – 0.60**	**0.60 – 0.75**	**0.75 – 1.00**
Speech Intelligibility	Bad	Poor	Fair	Good	Excellent

The STI parameter as described in this section is based on the definition given in IEC 60268-16 Ed. 4.0. The ISO 3382-3 standard for open plan offices [7] uses a slightly different definition of the STI, in that is does not take auditory masking and the hearing threshold into account. Also, the STI in ISO 3382-3 is based on a different (unisex) speech spectrum.

DIRAC contains a separate “STI for ISO 3382-3” parameter.

Room Acoustics Speech Transmission Index (RASTI)

The RASTI is a simplified version of the STI, intended to approach the STI under typical room acoustical conditions. The RASTI was originally developed to reduce the time required to perform the measurement (using modulated noise) and the time to compute the final result. In order to obtain correct RASTI values, in addition to the requirements for the STI method, the overall system frequency response must be uniform from the 125 Hz through the 8 kHz octave band, the background noise must be smooth in time and frequency, the space must be substantially free of discrete echoes and the reverberation time must not be strongly frequency dependent. For details, refer to [2].

The RASTI is calculated as the weighted sum of MTIs over the 500 and 2000 Hz octave bands, where the MTI values are derived from MTF values over 4 and 5 different modulation frequencies respectively. See Table 4.

Table 4
RASTI modulation reduction matrix

Modulation Frequency F [Hz]	**Octave Band Frequency [Hz]**	
	500	**2k**
0.7		●
1.0	●	
1.4		●
2.0	●	
2.8		●
4.0	●	
5.6		●
8.0	●	
11.2		●

In DIRAC, where the STI is measured through impulse responses rather than modulated noise, the RASTI has no advantage over the STI with respect to measurement or computation time. Nevertheless, the RASTI requires a smaller measurement system bandwidth than the STI, and it can be used for survey measurements in most practical room acoustical situations. The RASTI parameter has become obsolete in IEC 60268-16 Ed. 4.0. It should therefore no longer be used.

Speech Transmission Index for Telecommunication Systems (STITEL)

The STITEL is another simplified version of the STI, also meant to reduce measurement and calculation time. In order for the STITEL values to approach the corresponding STI values, a number of measurement conditions have to be met, which are typical for telecommunication systems.

The STITEL uses the same octave bands as the STI, but in each band only one octave band specific modulation frequency is used. See Table 5.

Table 5
STITEL modulation reduction

Modulation Frequency F [Hz]	Octave Band Frequency [Hz]						
	125	**250**	**500**	**1k**	**2k**	**4k**	**8k**
0.71			●				
1.12	●						
1.78						●	
2.83				●			
4.53							●
6.97					●		
11.33		●					

In DIRAC, where the STI is measured through impulse responses rather than modulated noise, the STITEL has no advantage over the STI with respect to measurement or computation time, but it is meant for comparability with other measured STITEL values.

Speech Transmission Index for PA Systems (STIPA)

The STIPA is a simplified version of the STI, intended to emulate STI under conditions, typical for public address systems. STIPA was developed to reduce the time required to perform a measurement (using modulated noise), and the time to compute the final result. In order for STIPA values to approximate corresponding STI values, several measurement conditions have to be met, which are typical for public address systems.

Table 6
STIPA modulation reduction matrix

Modulation Frequency F [Hz]	Octave Band Frequency [Hz]						
	125	**250**	**500**	**1k**	**2k**	**4k**	**8k**
0.63			●				
0.80						●	
1.00		●					
1.25					●		
1.60	●						
2.00				●			
2.50							●
3.15			●				
4.00						●	
5.00		●					
6.25					●		
8.00	●						
10.00				●			
12.50							●

In DIRAC, where STI is measured through impulse responses rather than modulated noise, STIPA has no advantages over STI with respect to measurement or computation time, but is meant for comparisons with other measured STIPA values.

Percentage Articulation Loss of Consonants (% ALC)

The % ALC (also called % ALcons) is originally based on the reception of words by listeners. In DIRAC, the % ALC is derived from the STI through a widely used approximation formula by Farrel Becker* [8]:

$$\% \text{ ALC} = 170.5405\, e^{-5.419(\text{STI})}$$

* The editor of [6] notes on p.81: This equation, referred to as Farrel Becker equation, is often used to relate STI to ALcons scores. It appears that the source of this equation is not documented in open literature. However, a remarkable correspondence is observed with the empirical data reported (in a figure rather than as an equation) by Houtgast et al. [1]. It seems reasonable to assume that the equation was either obtained through similar experiments, or derived from the data reported by Houtgast et al.

The same formula is used with RASTI, STITEL and STIPA. The % ALC values normally range from 0 (corresponding to an excellent speech intelligibility) to 100 (corresponding to an extremely bad speech intelligibility), but the % ALC value calculated from the above-mentioned approximation formula, will exceed 100 at a very low STI.

The % ALC in DIRAC is mainly meant for comparison with other calculated or measured % ALC values.

Signal-to-Noise Ratio (SNR)

The SNR is defined as the logarithmic ratio of the signal level and the noise level, and is therefore related to signals rather than systems. In DIRAC, the SNR is the ratio of speech and background noise signal levels as derived from the impulse response, or from the impulse response and a separate background noise measurement (intermittent measurement). This is not necessarily the same as the apparent SNR used in the speech intelligibility calculations. In certain measurements scenarios (described later in this document), the measured signal levels are modified to line up with prescribed speech spectra.

Early Decay Time EDT

Because the EDT relates more than the other reverberation parameters to the initial and highest level part of the decaying energy, it also relates most to modulation reduction. The EDT is derived from the decay curve section between 0 dB and 10 dB below the initial level. From the corresponding slope, the EDT is calculated as the time to reach –60 dB.

Measuring the Speech Intelligibility

Measurement Techniques

There are two factors that determine the speech intelligibility. The background noise, or rather the signal to noise ratio SNR, and the acoustics of the room (e.g. the sound reflections in the room). On page 16 the impact of the SNR on speech intelligibility is detailed, which leads to the important 15 dB SNR criterion:

> *The background noise is negligible if the SNR exceeds 15 dB in each relevant octave frequency band and the STI does not exceed 0.8.*

In practice, STI values rarely exceed 0.8 and therefore, in most practical cases it will be sufficient to meet only the SNR > 15 dB condition, in which case a background noise measurement is not required and the STI can be calculated from any impulse response measurement. In most interesting cases however the SNR is lower than 15 dB and the background noise level is relevant. *In the remainder of this document it is assumed the 15 dB criterion is not met, unless stated otherwise*.

DIRAC derives the acoustics of a room from a measured impulse response. Rife [4] has found that (under specific conditions) the SNR can also be calculated from the measured impulse response. This means that both factors influencing the speech intelligibility can be measured simultaneously in principle. The specific conditions that need to be fulfilled for an accurate determination of the SNR from an impulse response are low distortion in the measurement chain and time-invariant acoustics (no air movement, or people walking around), resulting in system induced noise that is much lower than the background noise. Also, pre-averaging cannot be used as this would artificially increase the SNR.

However, it is also possible to measure both factors independently. In fact, measuring the room acoustics using an impulse response without the presence of (significant) background noise gives much better results, as you can use pre-averaging, longer sequence lengths etc. to improve the quality (INR) of the impulse response. Also, it is not always possible or even desirable to measure the impulse response in the presence of background noise.

Measuring the background noise is also easier (an more accurate) through a direct level measurement instead of indirectly through an impulse response measurement.

Sound Sources

Mouth-directional loudspeaker sound sources

Mouth-directional loudspeaker sound sources, in short mouth simulators, have a directivity similar to a human mouth. This directivity is relevant for the speech intelligibility, in that speech intelligibility is highest on the axis of the source. An artificial mouth with directivity characteristics according to ITU-T P.51 [5] is preferred. A small high quality loudspeaker with diameter not exceeding 100 mm is also usable.

A mouth simulator is normally used in the situation of an unamplified talker or in a situation with a sound system equipped with a close-talking microphone.

Echo Speech Source Type 4720 was specifically designed for speech intelligibility measurements with DIRAC. It features the proper directivity and fully calibrated sound levels and sound spectrum. The use of the Echo with DIRAC is described in later sections of this document.

Omnidirectional loudspeaker sound sources

Omnidirectional sound sources are required for ISO 3382-3 compliant measurements. The source must fulfill the requirements set out in ISO 3382-1, and must be used at a height of 1.2 m. OmniSource Type 4295 and the OmniPower Type 4292-L are both omnidirectional sound sources that can be used for open plan office measurements.

Stimuli

Sweep versus MLS

For impulse response measurements any broadband signal can be used, but the most common are MLS and sweep. The exponential sweep (e-sweep) in particular has a number of properties making it very useful for impulse response measurements [9]. The disadvantage of e-sweeps (or sweeps in general) is that they sound much more intrusive than MLS signals. There are occasions where the use of sweeps is simply not allowed. Sweeps are also more difficult to handle for sound sources. The MLS is therefor still a very useful alternative.

Intermittent stimulus

To facilitate the independent measurement of impulse response (IR) and background noise, DIRAC provides the intermittent stimulus. This stimulus consists of a standard deterministic MLS or sweep stimulus followed by a period of silence. Both parts are treated as a single measurement, and the resulting impulse response and background noise measurement are combined in a single .wav file, with the IR in the first and the background noise in the second channel. The intermittent stimulus also circumvents Rife's condition regarding the artificial increase of the SNR when using pre-averaging.

Speech levels and spectra

Speech intelligibility measurements must be carried out using test signals that resemble human speech in level and spectrum in order to capture the correct SNR values in each frequency band. IEC 60268-16 specifies male and female speech spectra for use in STI, STIPA and STITEL measurements. The STIPA is only measured for a male speech spectrum. The A-weighted level of the speech signal should be 60dB at 1m from the source, and the octave bands should have the relative levels as indicated in Table 6.

The preferred method of measuring the speech intelligibility using DIRAC is therefor through separate measurements of the impulse response and the background noise. These can then be combined when calculating the speech intelligibility. This also makes it possible to use a single impulse response measurement to study different scenarios with different SNR values.

Table 7
Relative octave band levels [dB]

Octave Band [Hz]	125	250	500	1000	2000	4000	8000
Male	2.9	2.9	–0.8	–6.8	–12.8	–18.8	–24.8
Female	–	5.3	–1.9	–9.1	–15.8	–16.7	–18.0

For RASTI measurements, where only 2 octave bands are relevant, the relative levels are in Table 8.

Table 8
Relative octave band levels for the RASTI test

Octave Band [Hz]	500	2000
Level [dB]	–1.0	–10.0

ISO 3382-3 uses a spectrum derived from averaged male and female spectra [10]. The total A-weighted level is 57.4 dB at 1 m from the (omnidirectional) source with the levels for each octave band given in Table 9.

Table 9
Sound pressure levels of speech for ISO 3382-3

Octave Band [Hz]	125	250	500	1000	2000	4000	8000
Omnisource Level	49.9	54.3	58.0	52.0	44.8	38.8	33.5

Although it is possible to create signals with the correct spectrum in DIRAC, in most practical measurement scenarios an arbitrary but known source spectrum is used for the measurements and the measured SNR values are then recalculated for the desired speech spectrum.

Note that the frequency response of the source has to be taken into account when attempting to generate signals with a predefined spectrum. Although in principle this can be done using an equaliser or using a shaping filter created with DIRAC, it is much easier to use a calibrated source.

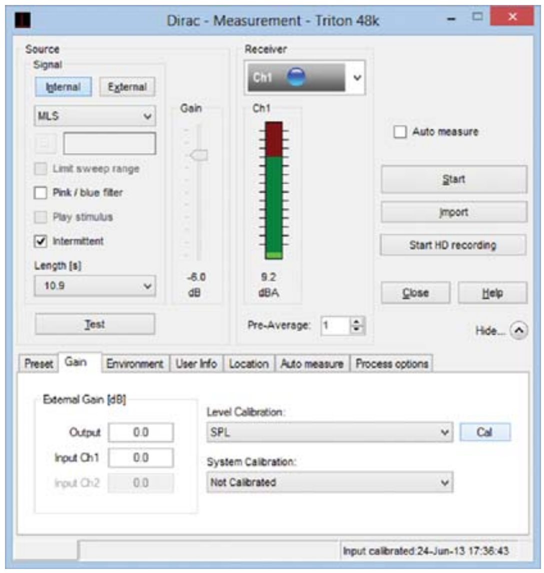
When the signal is injected directly in an existing PA system, no equalisation is necessary, as the response of the PA system always influences the speech signal, and is part of the system that needs to be measured. Note however that the stimulus still needs to be speech shaped if the 15 dB criterion is not met.

Level Calibration

In order to measure absolute sound levels in DIRAC, an input level calibration must be performed. To perform an input level calibration, follow these steps:

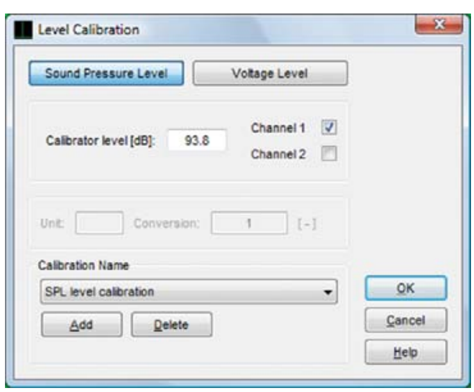
1. Click , the **Measure** button, and select the *Gain* tab in the lower half of the measurement window.

Fig. 2
Level calibration



2. Connect the omnidirectional microphone to the sound device input you want to use.
3. Insert the microphone into a single tone sound calibrator and switch the calibrator on.
4. Click the **Cal** button and wait until the measurement has finished. Make sure that the calibration signal did not stop before the measurement finished.
5. In the Level Calibration dialog box, fill in the sound calibrator level given by the manufacturer, select the channel on which you recorded the calibrator tone, and then click **OK**. Now the input level is calibrated for the selected channel.

Fig. 3
The level calibration dialog



6. Verify the calibration by switching the calibrator on again and check the value (in dBA) displayed beneath the level meter. It should display the calibrator level.

Note:
- The most recent input level calibration date is displayed in the lower right-hand corner of the Measurement window
- You can clear the calibration results by clicking the **Clear** button. This will also clear the input calibration notification in the Measurement window
- If you calibrate channel 1, while channel 2 has not yet been calibrated since the last calibration reset, channel 2 will adopt the calibration result of channel 1. This is not as accurate as calibrating channel 2 separately (which is recommended), but with two microphones of the same type, this is probably better than not calibrating channel 2 at all. Channel 1 will never adopt the calibration result of channel 2

- The input level calibration is valid only for the microphone and sound device setup used with the calibration
- The procedure is described for one microphone. You can also calibrate two microphones simultaneously, using a sound intensity calibrator
- Input level calibration is useful only if the sound device input frequency characteristic is flat within ±1 dB over 88 Hz through 11.3 kHz
- It is recommended to repeat the input level calibration before each speech intelligibility measurement session

System Calibration

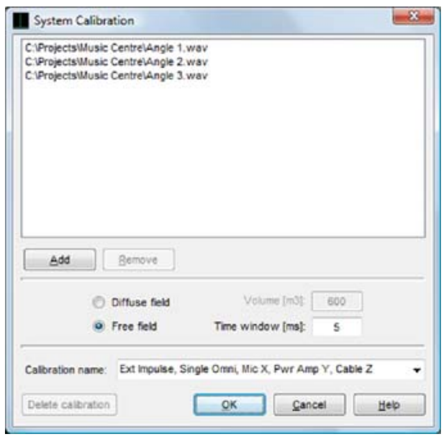


Fig. 4
The system calibration dialog

The sound source must deliver the speech signal at a prescribed sound level. For most speech intelligibility measurements one can use Echo Speech Source Type 4720. DIRAC is equipped to recognize the (fully calibrated) Echo signals making the Echo-DIRAC combination the easiest option for accurate speech intelligibility measurements.

ISO 3382-3 compliant measurements require the use of an omnidirectional sound source that needs to be calibrated. In DIRAC this is handled through a system calibration. The system calibration characterizes the complete measurement chain, including amplifiers and microphones. The calibration is best performed in a reverberation room, but a free-field calibration is also possible.

To perform the system calibration, follow these steps:

1. Set up the measurement window as you would for the speech intelligibility measurement. Make sure you select the same stimulus and receiver type, but do not select an intermittent stimulus. For instance, if you want to perform ISO 3382-3 measurements, load the ISO 3382-3 measurement preset, but uncheck the *Intermittent* option.
2. Make a note of the gain settings on you amplifier(s). These settings will need to be the same during the final measurement, or any changes will have to be entered on the *Gain* tab.
3. Perform impulse response measurements in a reverberation chamber or in a free-field situation (for instance on a large stage). In a reverberation chamber measure in at least three microphone positions. In a free field, rotate the source over 5, 7 or 9 angles of 360/*n* degrees.
4. Choose **Calibrate System…** from the **Setup** menu.
5. Click **Add** and select the calibration measurements.
6. Select *Diffuse field* and enter a *Volume* for a reverberation chamber, otherwise select *Free field* and enter an appropriate *Time window* to remove reflections from the signal.
7. Enter a suitable *Calibration name* or description and click **OK**.
8. In the measurement window, select the created system calibration on the *Gain* tab before starting the speech intelligibility measurements.

Note:
- You can perform the system calibration after the actual measurements, and apply the calibration to the measurements using the File Properties dialog. Through the Project window you can apply the system calibration to multiple files in a single operation
- Repeat the system calibration for each measurement configuration
- The free-field time window should be short enough to remove unwanted reflections from the response, yet long enough to get accurate results at low frequencies. Inspect the impulse responses visually to find a compromise value

Measurement Scenarios

Echo Speech Source Measurements (IEC 60268-16)

The Echo Speech Source is delivered with software. After installation on a PC where a copy of DIRAC is installed, DIRAC will provide the option to select the Echo signal as a stimulus, and it contains an Echo preset. To perform an Echo measurement, follow these steps:

1. Perform a level calibration of the microphone.
2. Select the Echo preset on the *Preset* tab of the measurement window.

3. If the Echo is positioned in front of a microphone connected to a PA system, start the speech signal and set the gain on the PA system such that the speech level appears normal.
4. Turn on the Echo Speech source and start the MLS signal.
5. Click **Start** in the DIRAC measurement window.

DIRAC will automatically recognize the continuous or intermittent signal and whether it is played at a normal or raised level.

Use the intermittent signal where the reverberation time is relatively short and the background noise has a significant impact. As a rule of thumb, use the intermittent signal when RT × SNR < 120. Otherwise use the continuous signal. Note that you can use pre-averaging with the intermittent stimulus to increase the INR.

The raised Echo level should only be used when the background noise is very high and pre-averaging does not get the INR over 20 dB for the normal signal level.

Open Plan Office Measurements (ISO 3382-3)

DIRAC contains a preset for ISO 3382-3 measurements. The procedure for open plan office measurements is as follows:

1. Perform a system calibration with the measurement equipment (you only need to do this once).
2. Perform a level calibration.
3. Select the ISO 3382-3 preset on the *Preset* tab of the measurement window.
4. The preset uses a pink filtered MLS stimulus. Change this to an e-sweep when desired, but note that the system calibration must be performed with the same stimulus.
5. Use the **Test** button to find the correct output gain. Change the amplifier gains if necessary, but be sure to enter the gain differences on the *Gain* tab.
6. Click the **Start** button in the DIRAC measurement window.
7. Repeat the measurement on 6 to 10 workstation positions along a line.
8. Analyse the impulse responses.

The ISO 3382-3 preset uses an intermittent stimulus. It is possible to set a higher pre-average value with this measurement to increase the INR if necessary.

Other Speech Intelligibility Measurements

When the 15 dB criterion (see Measurement techniques) is met, a speech intelligibility measurement can be performed with this procedure:

1. Connect the excitation signal to a mouth simulator at the talker position or to the sound system input.
2. Connect an omnidirectional microphone at a listener position to sound device input 1.
3. In DIRAC, select an excitation signal (e-sweep or MLS are common choices).
4. Set a capture length exceeding 2 s and the estimated reverberation time. Note that a longer capture length generally results in a better INR.
5. Select a Pre-average value. Higher pre-average values result in a higher INR.
6. Measure and save the impulse response.
7. If applicable, repeat the measurement with the receiver at different listener positions.
8. Analyse the impulse responses.

When the 15 dB criterion is not met, measuring an impulse response in the presence of significant background noise to determine the speech intelligibility can be difficult without a calibrated source. DIRAC demands an INR higher than 20 dB to be able to reliably calculate the SNR. At the same time, a low SNR value during the measurement is detrimental to the quality of the impulse response and hence the INR. To add to these difficulties, the method devised by Rife to calculate the SNR from the impulse response does not give the correct results when the SNR is increased artificially through pre-averaging.

1. Calibrate the input level.
2. Use an equalizer and a pink noise signal to flatten the response of a mouth simulator when applicable. You can also use the Shaping Filter Designer (SFD) in DIRAC to create a filter with which to shape the stimulus for a specific sound source.
3. Generate an MLS or Lin-Sweep stimulus (Edit – Generate) and filter it with an appropriate filter (Edit – Filter). Use the resulting signal as a User defined stimulus. Select the filter depending on the parameter that needs to be calculated (see Table 10).
4. Set the output level to a practical value or 60 dB(A).
5. Measure the IR with a pre-average value of 1 for a non-intermittent measurement, use a higher pre-average value if necessary for intermittent measurements.

Table 10
Output filter versus speech intelligibility parameter

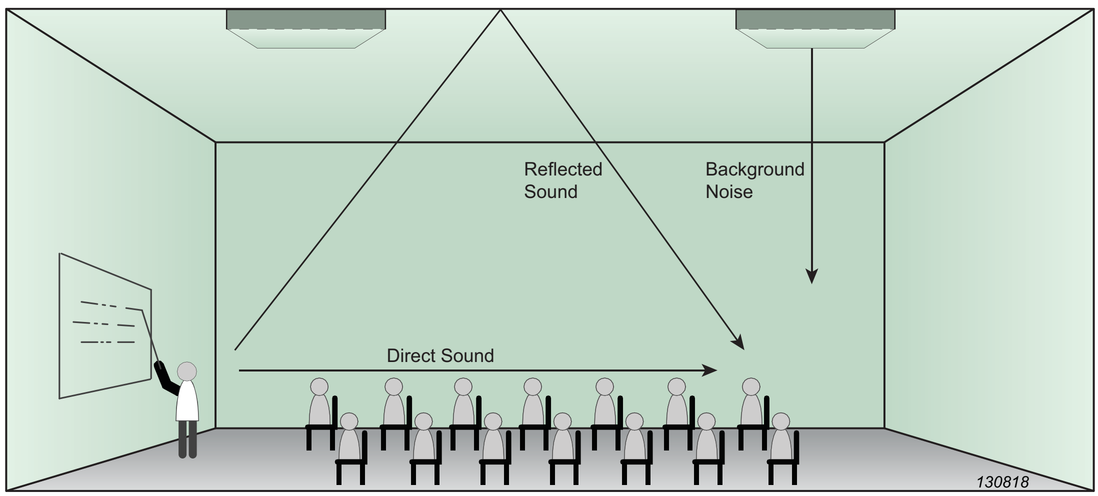
Speech Intelligibility Parameter to be Measured	Filter
STI Male or corresponding % ALC value	Male
STI Female or corresponding % ALC value	Female
RASTI or corresponding % ALC value	RASTI
STIPA or corresponding % ALC value	Male

Note that it is possible to use either internally or externally generated stimuli. When using an external stimulus, the stimulus type, source filter and sequence length of this external signal need to conform to the settings in the measurement dialog. You can find a number of speech filtered MLS and Lin-Sweep stimuli on the DIRAC CD.

At this point it is important to note that the INR in all relevant octave bands of the measured impulse response should be higher than 20 dB. This is required for DIRAC to be able to calculate the SNR and many other parameters. In most situations this requirement can easily be met by a combination of techniques such as longer capture lengths, higher pre-average values and higher stimulus output levels.

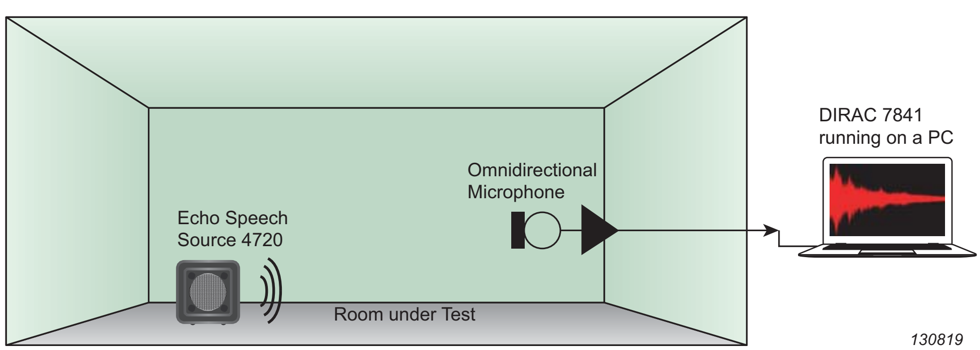
Situation 1: Talker and Listeners in Same Room without Sound System

Fig. 5
Room without sound system



The listeners receive direct sound from the talker, reflected sound, such as reverberation or echoes, and background noise, for instance from an HVAC system. The direct sound contributes positively to the speech intelligibility. Reflected sound may contribute positively (e.g., via the front board) or negatively to the speech intelligibility. Background noise contributes negatively to the speech intelligibility.

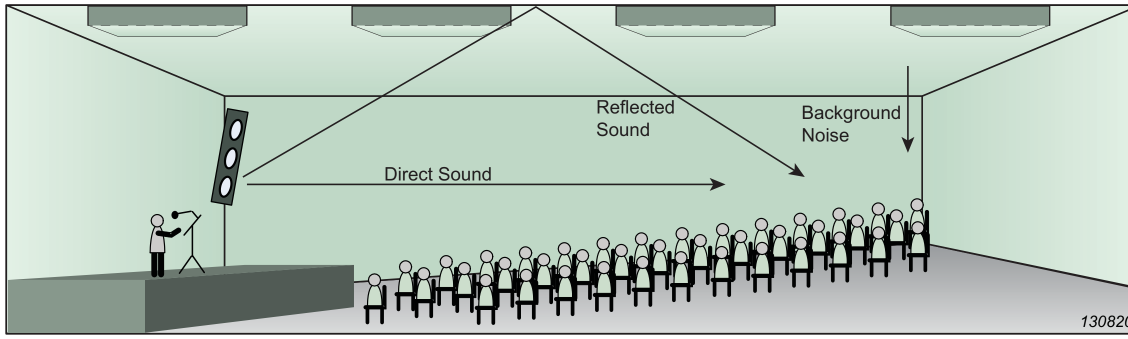
Fig. 6
Measurement setup in room without sound system



Situation 2: Talker and Listeners in Same Room with Sound System

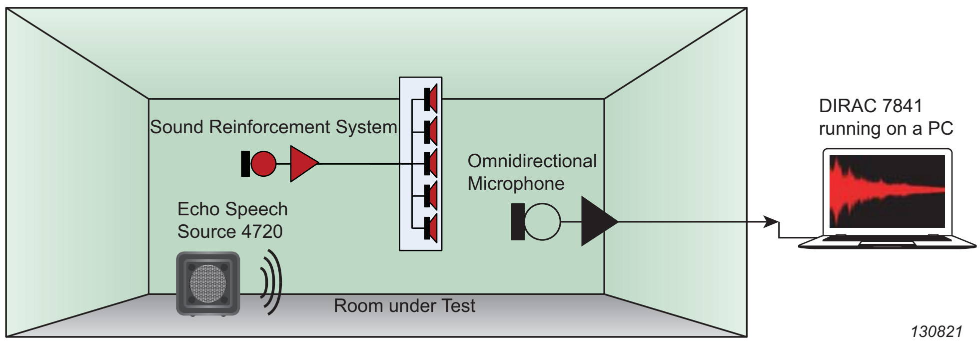
Fig. 7 shows a talker and listeners in a room with sound reinforcement system. Primarily, the listeners receive direct sound from the sound reinforcement system, reflected sound, such as reverberation or echoes, and background noise, for instance from an HVAC system. There are also some secondary effects. There is direct sound coming from the talker, which is a low level delayed version of the primary direct sound arriving at the listeners. Furthermore, the sound system microphone not only picks up the direct sound from the talker, but also reflections from the lectern, fed-back sound and background noise.

Fig. 7
Room with sound reinforcement system



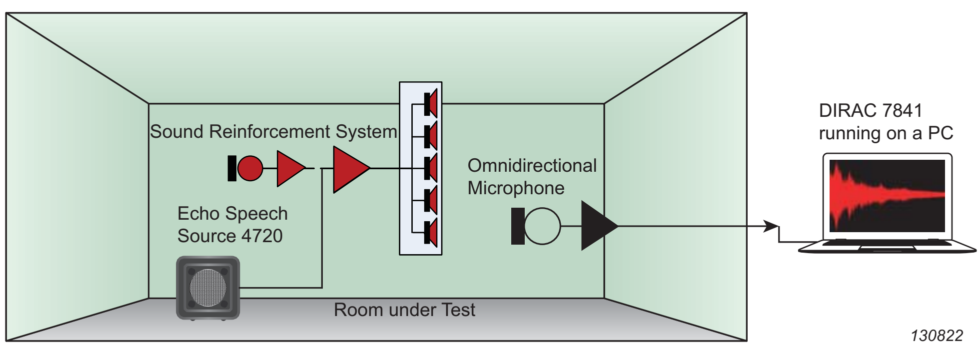
In this situation, there are two ways to measure the speech intelligibility. To include the sound system microphone characteristic and all secondary effects, use a mouth simulator.

Fig. 8
Measurement in room with sound system



Otherwise, the excitation signal can be injected directly into the sound system.

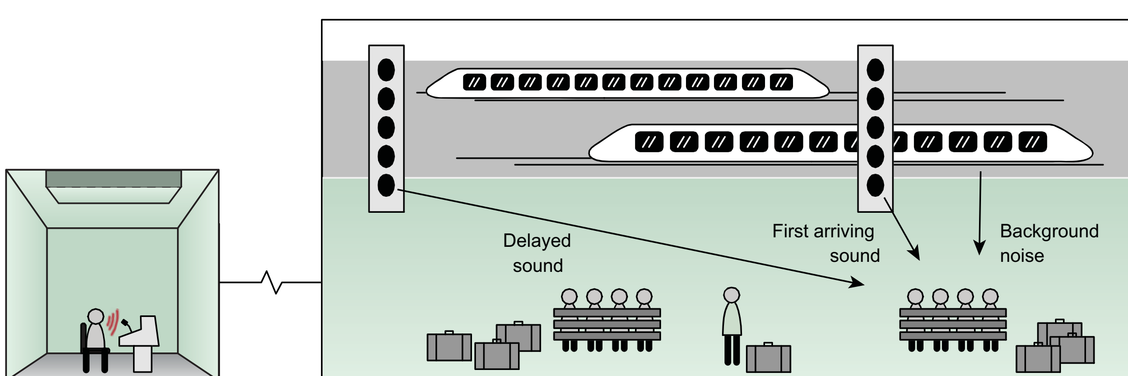
Fig. 9
Measurement setup using direct injection, in room with sound system



Situation 3: Talker and Listeners in Different Areas

Fig. 10 shows a talker and listeners in different areas connected through a long-distance sound system. Primarily, the listeners receive direct sound from a nearby loudspeaker array, delayed direct sound from a loudspeaker array further away, reflected sound, such as reverberation or echoes, and background noise, for instance from the environment. There are also some secondary effects. The sound system microphone not only picks up the direct sound from the talker, but also reflections and background noise from within the talker's room.

Fig. 10
Separate talker and listener areas with long distance sound system



Also in this situation, there are two ways to measure the speech intelligibility. To include the sound system microphone characteristic and all secondary effects, use a mouth simulator.

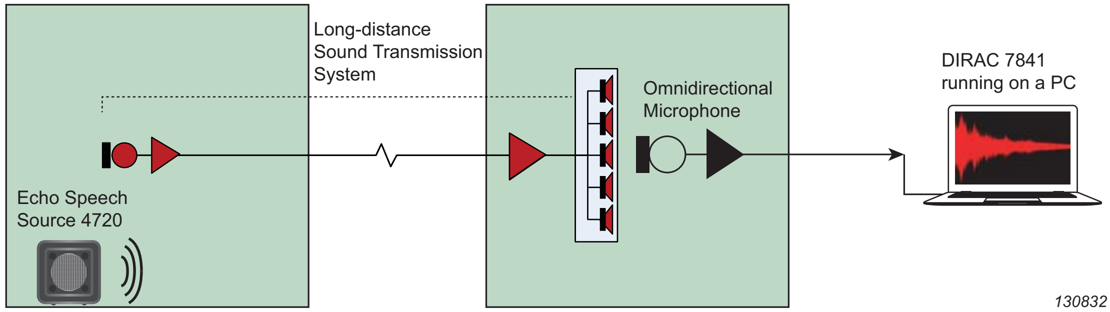


Fig. 11
Open loop measurement setup with Echo Speech Source

Otherwise, you can inject the excitation signal directly into the sound system.

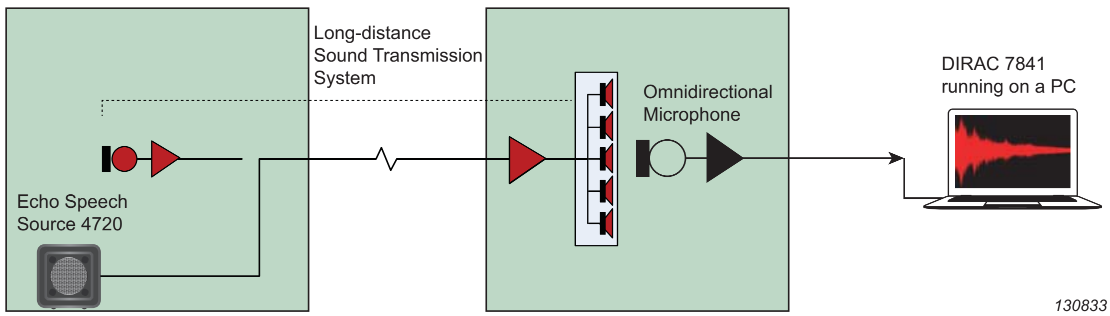


Fig. 12
Open loop measurement setup using direct injection

Measurement Analysis

Impulse Response Quality

It is important to check the quality of the measured impulse responses after each measurement, as low quality impulse responses can lead to erroneous results. There are 3 properties of an impulse response that can be checked easily and will give a good impression of the quality and usability of the impulse response.

The first thing to check is the Impulse to Noise Ratio (INR). The INR can be interpreted as the decay range of the impulse response. In general, the INR should be as high as possible. A minimum of 20 dB is barely acceptable. Values over 40 dB should be obtainable. Note that the INR can be improved by using higher signal levels, longer capture lengths and/or pre-averaging.

Another thing to check is whether the onset of the impulse response is detected correctly by DIRAC. The start of the impulse response is indicated by a vertical red marker in the impulse response view.

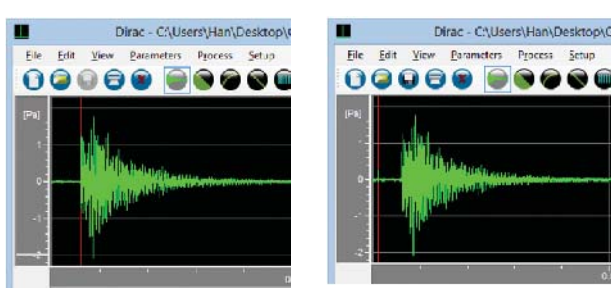


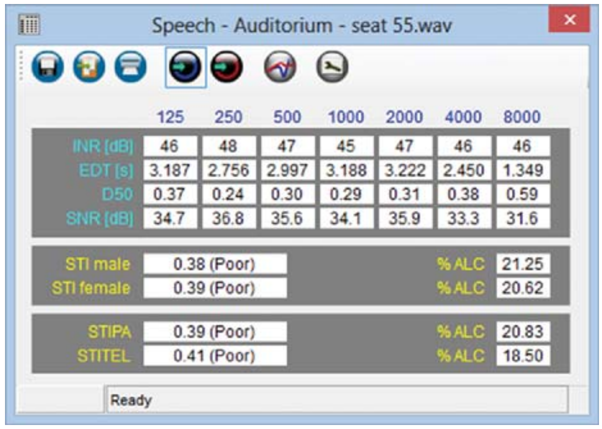
Fig. 13
Left: Correct detection of IR onset.
Right: Spurious peak causes erroneous detection

A final impulse response feature that must be verified is that the capture length is long enough given the impulse response. As a rule of thumb, the capture length should be twice the reverberation time. Another way to look at this is by seeing that the point where the impulse response decays into the background noise should be well within the measurement length. The measurement should clearly display a flat tail where the background noise dominates over the impulse response.

Parameter Tables and Graphs

The Speech table, which can be opened from the Parameters menu on the Impulse Response window, lists the speech intelligibility parameters for a given impulse response.

Fig. 14
The Speech table containing the calculated speech parameters

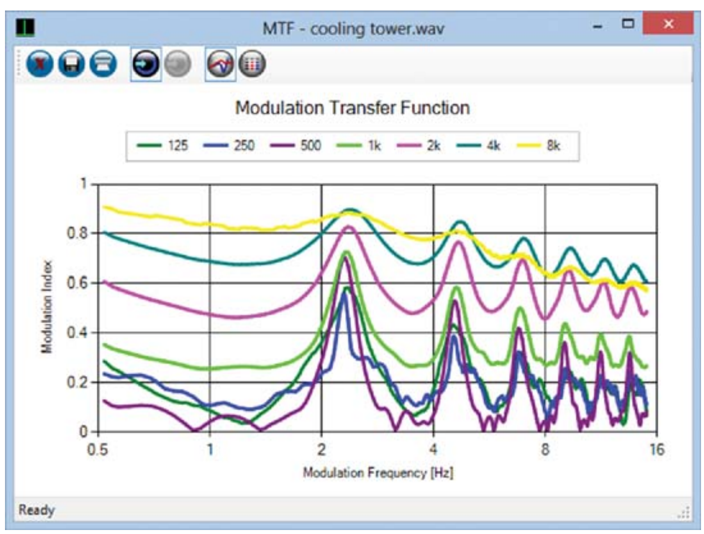


Note that the SNR displayed in the speech table indicates the SNR as measured at the microphone position. This is not necessarily the SNR used in the speech intelligibility calculations, as DIRAC may modify the signal levels for a specific speech spectrum and a normalized source level.

From the speech table you can quickly access the file properties using , the **Properties** button, where you can modify the signal and noise levels (see page 16).

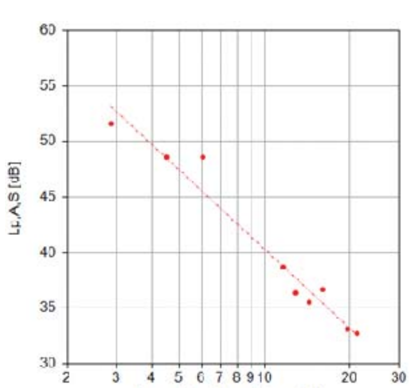
The MTF graph is displayed by using the MTF window button, , allowing you to study the causes of the modulation reduction (see page 3).

Fig. 15
MTF graph of cooling tower



Note that a continuous MTF graph as displayed above can only be calculated for impulse response measurements because modulated noise measurements only contain a limited set of discrete modulation frequencies.

Fig. 16
Regression line through a number of measurements



Individual parameters and parameter statistics over multiple files can be displayed in the Parameter Graph window. It will show you the average, minimum and maximum values, and their standard deviation. You can also plot the parameters versus the source-receiver distance and calculate parameters based on regression lines through these graphs such as the $L_{p,A,S,4}$, $D_{2,S}$, r_D and r_P parameters defined in ISO 3382-3.

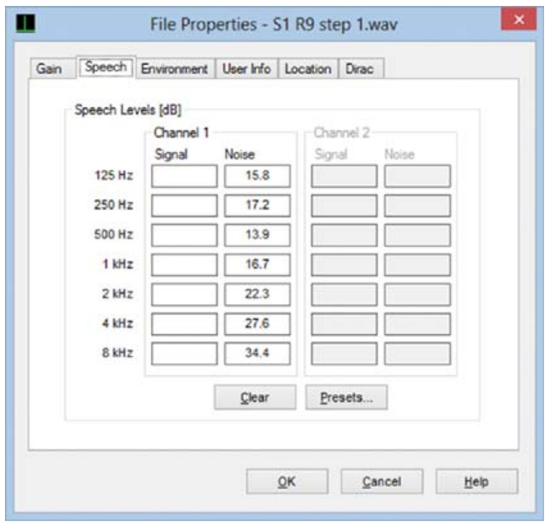
Changing Signal and Noise Level Values

It is often desirable to know what the speech intelligibility would be with signal or noise levels that differ from those at the time of the impulse response measurement. DIRAC provides several ways to investigate the resulting speech intelligibility when speech or noise signal levels are changed.

When the impulse response is measured with little background noise present, it is possible to add noise to the measurement. This can be synthetic noise (menu item: Process - Mix - Noise), or a recording of actual noise (menu item: Process - Mix - File).

In the File Properties window one can set the speech signal and noise values numerically. To investigate the effect of raised speech or noise levels, one can use the Speech and Noise parameter values as a starting point.

Fig. 17
Changing noise levels



Note that you can enter either noise levels, speech levels or both. DIRAC will use whatever was entered in the File Properties dialog, and use actual values from the measurement where no value is entered.

In the parameter graph window you can edit the file properties of multiple files at once. Use the Edit menu to change the properties of all files that have a checkmark set.

The Impact of SNR on Speech Intelligibility

If the background noise level is negligible, there is no need to measure the SNR and incorporate it in the speech intelligibility calculations. Therefore it may be worthwhile to take a closer look at the impact of the SNR on the speech intelligibility in practical situations, and formulate the exact condition under which the SNR is relevant.

The background noise is assumed to be negligible if its presence results in a STI decrease of less than 5% of the STI without background noise. To figure out what the consequence is for the allowed SNR, we write the MTF as a product of 2 modulation reduction factors, $m_0(k,F)$ due to system properties (reverberation, echoes) and $m_{SNR}(k,F)$ due to background noise:

$$MTF(k,F) = m_0(k,F) \cdot m_{SNR}(k) = \frac{m_0(k,F)}{1 + 10^{-\frac{SNR(k)}{10}}} \quad (1)$$

where k is the octave number. By definition, the effective SNR, $SNR_{eff}(k,F)$ relates to $MTF(k,F)$ as does the SNR to m_{SNR}:

$$SNR_{eff}(k,F) = 10\log\left(\frac{MTF(k,F)}{1 - MTF(k,F)}\right) \quad (2)$$

Hence, SNR_{eff} equals SNR if the modulation is reduced by background noise only. For each k from a set of octave frequency bands, and each F from a set of modulation frequencies, $SNR_{eff}(k,F)$ is calculated, and clipped to ±15 dB, before being further processed to calculate the STI. The clipping operation reflects that SNR_{eff} values exceeding 15 dB cannot have any negative impact on the speech intelligibility, while SNR_{eff}

values lower than –15 dB cannot have any positive impact on the speech intelligibility. The clipped SNR_{eff} values are converted to transmission indices $TI(k,F)$ that range from 0 to 1, and each contribute to the STI.

$$TI(k, F) = \frac{SNR(k, F) + 15}{30} \tag{3}$$

For the final STI, the TI values are averaged over the modulation and octave band frequencies in a special way, thereby taking into account auditory masking and the absolute hearing threshold. However, to get an idea of the impact of the SNR on the STI, it is sufficient to evaluate TI for several values of m_0 and SNR. Fig. 19 shows the relative change of TI when going from a situation with $SNR = \infty$ to a situation with the given finite SNR:

$$\Delta TI = \frac{TI_{\text{SNR}} - TI_{\infty}}{TI_{\infty}} \cdot 100\% \tag{4}$$

If the relative changes in TI would also hold for the STI in case of equal weighting factors for each octave frequency band. Actually, the SNR values of the octave bands from 500 Hz through 4 kHz are the most significant.

From Fig. 18 and Fig. 19 it can be seen that as m_0 increases from 0.5, TI changes faster with m_0 (hence with SNR), until SNR_{eff} clips to 15 dB at $TI = 1$, resulting in a dip at $m_0 = 0.97$. A much steeper, yet practically irrelevant dip at $m_0 = 0.03$ is not shown.

Using Fig. 19 and Table 10, in which relevant octave frequency bands are defined, we now define the **15dB SNR criterion**:

> *The background noise is negligible if the SNR exceeds 15 dB in each relevant octave frequency band and the STI does not exceed 0.8.*

If only the condition of the SNR exceeding 15 dB is met, this could theoretically lead to an underestimation of the STI of 10%, namely if the SNR is close to 15 dB for all relevant octave frequency bands and the STI without background noise would be 0.97, resulting in a measured STI of 0.9. In practice however, STI values rarely exceed 0.8 and most likely, only a few SNR values, if any, will come close to 15 dB. Therefore, in most practical cases it will be sufficient to meet only the SNR > 15 dB condition.

Fig. 18
Transmission Index as a function of the modulation index when no background noise is present

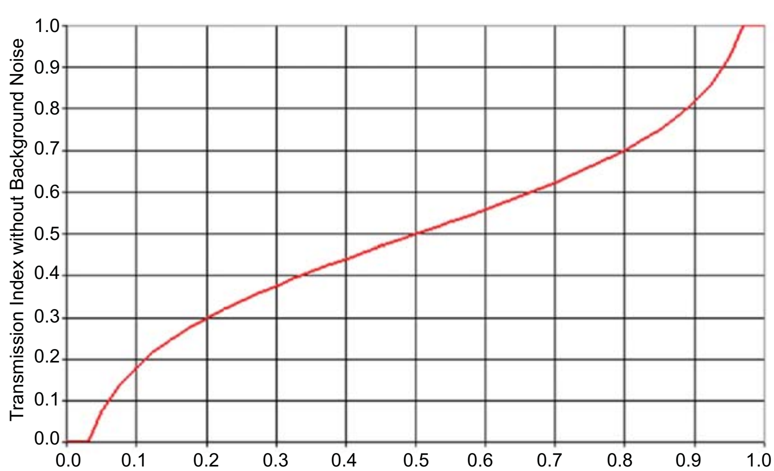


Fig. 19
Relative change of the Transmission Index as a function of the modulation index for different background noise levels

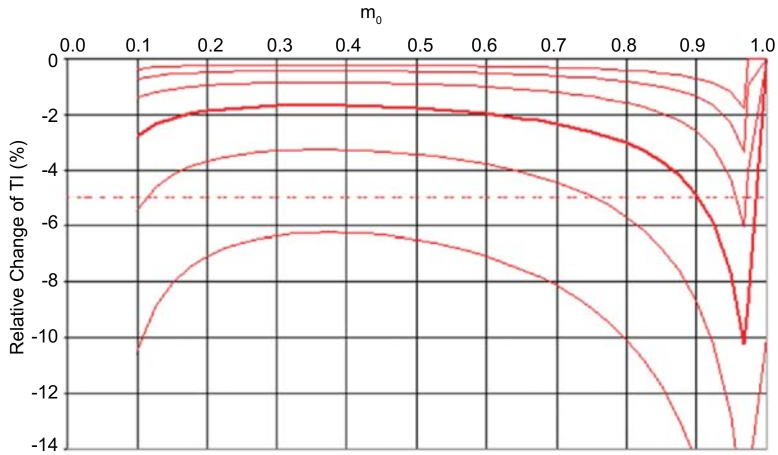


Table 11
Relevant octave frequency bands

Octave Frequency Band [Hz]	Source Excitation Signal		
	Male	Female	RASTI
125	•		
250	•	•	
500	•	•	•
1k	•	•	
2k	•	•	•
4k	•	•	
8k	•	•	

References

[1] T. Houtgast, H.J.M. Steeneken and R. Plomp, “Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. I. General Room Acoustics,” *Acustica 46, 60 – 72 (1980)*

[2] IEC 60268-16 Edition 4.0 (2011) Sound system equipment. Part 16: Objective rating of speech intelligibility by speech transmission index

[3] M.R. Schroeder, “Modulation Transfer Functions: Definition and Measurement,” *Acustica 49, 179 – 182 (1981)*

[4] D.D. Rife, “Modulation Transfer Function Measurement with Maximum-Length Sequences”, *J. Audio Eng. Soc. 40, 779 – 790 (1992)*

[5] ITU-T Recommendation P.51 (03/93), “Telephone Transmission Quality Objective Measuring Apparatus – Artificial Mouth”

[6] T. Houtgast, H.J.M. Steeneken et al., “Past, present and future of the Speech Transmission Index”, *TNO Human Factors, Soesterberg, The Netherlands, © 2002*, ISBN 90-76702-02-0

[7] ISO 3382 (2012), “Acoustics – Measurement of room acoustics parameters – Part 3: Open plan offices”

[8] “A Do-it-Yourselfers Guide to Computing the Speech Transmission Index”, http://www.synaudcon.com/site/articles/a-do-it-yourselfers-guide-to-computing-the-speech-transmission-index/

[9] A. Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique”, *110th AES Convention, February 2000*

[10] “Guidance on computer prediction models to calculate the speech transmission index for BB93.” Version 1.0, *Department for Education and Skills, Schools Capital and Building Division, 2004*

BO 0521 – 11   2013-12

HEADQUARTERS: Brüel & Kjær Sound & Vibration Measurement A/S · DK-2850 Nærum · Denmark
Telephone: +45 7741 2000 · Fax: +45 4580 1405 · www.bksv.com · info@bksv.com

Local representatives and service organisations worldwide

Brüel & Kjær