ODEON APPLICATION NOTE – Calculation of Speech Transmission Index in rooms

JHR, February 2014

Scope

Sufficient acoustic quality of speech communication is very important in many different situations and places; classrooms in schools, auditoria, PA systems in public buildings, alert and warning systems in noisy environments etc.

Speech transmission index (STI) is an objective method for prediction and measurement of speech intelligibility. This is a guide how to use ODEON for calculating the speech transmission index in a room in accordance with the international standards IEC 60268-16 [1] and ISO 9921 [2].

This application note is not applicable for open-plan offices, which are dealt with in another ODEON application note[1].

This application note refers to ODEON version 12, all editions.

Method

Although IEC 60268-16 is a measurement method, the measurement procedure can be simulated in ODEON, thus providing a tool for the acoustical design of performance spaces and other rooms where speech communication is important, or for the design of a loudspeaker system in a room.

Two different types of cases will be discussed: Direct communication between two people in the same environment and public address (PA) systems where an electro-acoustic system is used to address a group of people.

The original method [1] is based on a special modulated speech-shaped signal and the measurement of modulation transfer functions in seven octave bands with 14 different modulation frequencies. However, an alternative and equally good method uses the impulse response as suggested by Schroeder [3]. This so-called indirect method is the one used in ODEON.

Terms and definitions

Ambient noise
Noise due to all other sound sources than the communication speech.

Background noise level
Sound pressure level in octave bands due to ambient noise, spatially averaged over the relevant listener area.

[1] http://www.odeon.dk/pdf/Application_note_ISO-3382-3_Offices.pdf

Public address system
PA
Electro-acoustic system used to address a group of people.

Reference speech level
Speech level equivalent to 60 dB(A) at 1 m distance in front of the talker's mouth.

Signal-to-noise ratio
SNR
Difference between the sound pressure level of the speech or test signal and the sound pressure level of the background noise.

Speech intelligibility
Rating of the proportion of speech that is understood.

Speech transmission index
STI
Objective method for prediction and measurement of speech intelligibility.

Rapid speech transmission index
RASTI
A simplified and approximate measurement method for rating speech transmission in auditoria, with or without sound systems.

NOTE 1: RASTI was defined in IEC 268-16:1988.

NOTE 2: RASTI is obsolete and should not be used any more.

Speech transmission index for PA
STIPA
A condensed and approximate version of the STI measurement method for PA systems.

NOTE: STIPA is only applied as a direct method, not relevant for the indirect method used in simulations.

STI (expected)
Value of STI calculated from the reverberation time and assuming an ideal exponential decay of energy.

NOTE: Due to direct sound and beneficial early reflections STI will normally be greater than STI (expected). So a comparison of the actual STI with STI(expected) can reveal the quality of the acoustic design in the room. If STI is greater than STI(expected) then the design is considered better that if the room was totally diffuse.

Tapping
Is the process of adjusting the electric power applied to a constant-voltage speaker system, by connecting various input taps to the step-down transformer. The adjustment of the electric power leads to an adjustment of the sound power emitted by the loudspeaker unit.

Vocal effort
Characterisation of the sound level emitted by a human speaker, quantified objectively by the A-weighted speech level at 1 m distance in front of the mouth and qualified subjectively by a description.

Voice alarm system
VA
Electro-acoustic system used to broadcast speech messages or warning signals in an emergency.

Sound sources

Three different types of sound source are considered in this application note, a human speaker, a sound reinforcement loudspeaker system and a constant-voltage speaker system, also called a high-voltage audio distribution system.

The case of a human speaker applies to class rooms, meeting rooms and small auditoria. The speech level and speech spectrum have to be chosen according to the preconditions for the room.

Cases where the sound source is one or more loudspeakers include auditoria, concert halls and stadiums. The constant-voltage speaker systems are typically applied in public areas, e.g. shopping centres, railway stations, airports. Other important applications are voice alarm (VA) systems in noisy working places, e.g. off-shore installations.

Human speakers

The important acoustical parameters for a sound source representing a human speaker are the speech level and the directivity, both given in octave bands. The source for generalised speech level data is the American Standard ANSI 3.5 [4], which has data representing four different levels of vocal effort; normal, raised, loud, and shouted, see Table 1.

The directivity of human speakers is available from a Canadian study [5]. In this research report several factors were investigated, and it is concluded that the directivity is not significantly affected by vocal effort, gender, or language (English or French).

Among the predefined sound sources in ODEON you will find two that are relevant to this application: BB93_NORMAL_NATURAL.SO8 and BB93_RAISED_NATURAL.SO8. The name of the source files reflect the fact that these sources are as defined in the calculation guidance to the British Building Bulletin 93 [6]. The data at the 125 Hz octave band are not available from ANSI 3.5 but are required for the STI calculations, and thus the speech level at 125 Hz has been chosen to be 6 dB lower than the 250 Hz level.

The original data for speech level are given as sound pressure levels (SPL) at a distance of 1 m in front of the speaker's lips in the free field. This kind of data is relevant for a communication system where a microphone is placed in front of the speaker. However, for room acoustic applications it is more useful to define the sound source by the radiated sound power. Consequently, these data have been calculated and are also shown in Table 1.

Table 1 – *Speech spectra in octave bands for different vocal effort. SPL at 1 m in front of the mouth from ANSI 3.5 [4] and the calculated corresponding sound power levels. The additional data at 63 and 125 Hz have been suggested in [6].*

Frequency, Hz	63	125	250	500	1000	2000	4000	8000	A-weighted
ANSI 3.5, SPL at 1 m, dB									
Normal			57,2	59,8	53,5	48,8	43,8	38,6	59,5
Raised			61,5	65,6	62,3	56,8	51,3	42,6	66,5
Loud			64,0	70,3	70,6	65,9	59,9	48,9	73,7
Shouted			65,0	74,7	79,8	75,8	68,9	58,2	82,3
Sound power levels, dB									
Normal	*45,0*	*55,0*	65,3	69,0	63,0	55,8	49,8	44,5	68,4
Raised	*48,0*	*59,0*	69,5	74,9	71,9	63,8	57,3	48,4	75,5
Loud	*52,0*	*63,0*	72,1	79,6	80,2	72,9	65,9	54,8	82,6
Shouted	*52,0*	*63,0*	73,1	84,0	89,3	82,4	74,9	64,1	91,0

Vocal effort

Which vocal effort to choose, depends on the circumstances. For application in schools the British guideline BB93 [7] specifies the *raised* voice effort when calculating STI for teacher to student communication and the *normal* voice effort for student to student communication.

ISO 9921 [2] has slightly different definitions and more levels of vocal effort, see Table 2.

Table 2 – Vocal effort of a male speaker and related A-weighted SPL at 1 m in front of the mouth [2].

Vocal effort	$L_{S,\,A,\,1\,m}$ dB
Relaxed	54
Normal	60
Raised	66
Loud	72
Very loud	78

In a noisy environment, the very loud or shouted vocal effort may be relevant. If SPL at 1 m exceeds 75 dB(A) (shouted), the speech quality is substantially reduced, making it more difficult to understand. For that case the normative Annex A of ISO 9921 [2] specifies that the STI calculation shall be made with a speech level that is reduced by:

$$\Delta L = 0{,}4\left(L_{S,A,1\,m} - 75\right)\text{dB} \quad \text{for} \quad L_{S,A,1\,m} > 75\ \text{dB}$$

So, if for instance the vocal effort of shouted is assumed, the spectrum shall be as in Table 1, but the overall A-weighted level should be reduced by 0,4 (82,3 – 75) = 2,9 dB. In this way the effective signal-to noise level is reduced, which leads to a somewhat lower result of the STI calculation, and thus the reduced speech quality is taken into account.

Loudspeakers - Input signal

For a loudspeaker system the acoustic chain starts with speech that reaches a microphone, followed by amplifiers and equalizers and one or more loudspeakers.

In relation to the STI parameter the input signal is supposed to represent speech and the input to the electro-acoustic system may be from a microphone or playback. The input signal may then depend on the gender (male or female speech), vocal effort, and distance from the mouth to the microphone. Speech spectra for male and female speech that should be used for the STI calculations are given in Table 3.

Table 3 – Speech spectra in octave bands relative to the A-weighted SPL for different gender. This data are from IEC 60268-16 [1], except the 125 Hz value for females, which has been suggested in ISO 9921 [2].NB, these speech spectra should only be applied to electro-acoustic systems.

Frequency, Hz	63	125	250	500	1000	2000	4000	8000	A-weighted
Males, dB	-	2,9	2,9	-0,8	-6,8	-12,8	-18,8	-24,8	0,0
Females, dB	-	*-4,4*	5,3	-1,9	-9,1	-15,8	-16,7	-18,0	0,0

Comparing the IEC speech spectra in Table 3 with the ANSI speech spectra in Table 1, there are obvious differences. While the ANSI spectra apply to the far field (1 m or longer distances), the relative IEC spectra in Table 3 apply to close distances, typically a microphone distance of only 5 cm from the mouth. The origin of the IEC spectra is dull, but may go back to very early measurements at the Bell Laboratories in the 1930'ies or 1940'ies. The spectra show unexpected high values at the low frequencies (125 – 250 Hz), which may be explained by the so-called *proximity effect*[2].

The male speech is generally recommended to assess speech transmission channels [1, A.3.4]. The reason is that female speech is considered to be more intelligible than male speech, so the latter is a kind of "worst case". In addition to different speech spectra, there also applies gender-specific weighting factors, so e.g. the 2 and 4 kHz octave bands are given more weight for female speech compared to male speech. The parameters STI (male) and STI (female) are calculated by ODEON.

Loudspeakers for sound reinforcement - Output signal

The important acoustical parameters for a sound source representing a loudspeaker are the sensitivity, the frequency response and the directivity in octave bands. Loudspeakers for sound reinforcement are low impedance speaker systems (8-ohm speaker systems).

Loudspeaker types

Low impedance loudspeakers are divided into different types: *Passive loudspeakers*, for direct connection to a power amplifier; *active loudspeakers*, with equalization/processor/cross-over required before amplification, and *powered loudspeakers*, with built-in equalization/processor/amplification.

[2] Proximity effect is the increase in the low-frequency sensitivity of a microphone when the sound source is close to it. The proximity effect may be responsible for speech spectra showing emphasis in the low-frequency range.

Data for the acoustic simulation are best taken from the Common Loudspeaker Format (CLF) database. There are two file formats, either CLF1 in octave bands or CLF2 in 1/ 3 octave bands and with a finer angular resolution of the directivity patterns. Both file formats can be taken directly into the ODEON source editor. An example of the data information in a CLF1 file inside the CLF viewer is seen in Figure 1. The CLF viewer is free for download from this site: http://www.clfgroup.org/viewer.htm. The loudspeaker shown here will be applied later in example 2.

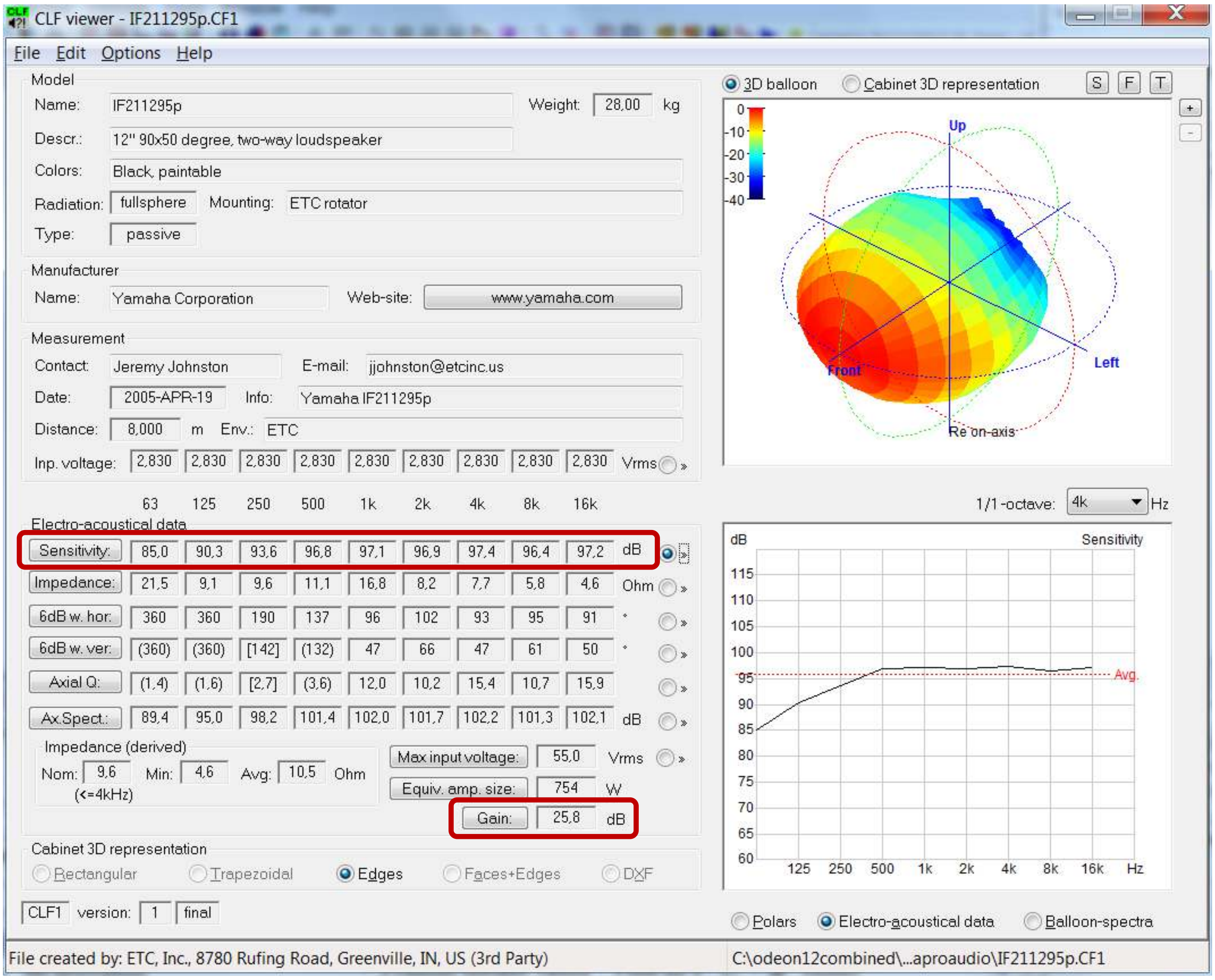

Figure 1 – Example of loudspeaker data in the CLF1 format. Marked in red: Sensitivity and maximum gain.

The maximum input is either in terms of electric power (in W) of as input voltage (in V rms). In either case the Gain (in dB) indicates the maximum setting for the loudspeaker. The sensitivity is the SPL in octave bands on axis at 1 m with a gain of 0 dB (typical reference 1 W or 2.83 V – according to the 8 Ohm loudspeaker convention). In the ODEON source editor menu the overall gain and the equaliser can be used to adjust the actual output level of the loudspeaker.

Constant-voltage loudspeakers - Output signal

The amplifier uses a transformer to step up the voltage of the audio signal to reduce the loss over the speaker cable, and the voltage is stepped back down at the loudspeaker. Typically used for PA and VA systems where high sound quality is not required. Commonly referred to as 70-volt speakers (in North

America) or as 100-volt system (Europe). When the loudspeaker data are in the correct format, the overall gain = 0 dB corresponds to the electrical input of 1 W. When the actual tapping is different from 1 W, the corresponding value of overall gain is calculated from:

$$\text{Overall Gain} = 10 \log\left(\frac{\text{tapping}}{1\ \text{W}}\right)\ \text{dB}$$

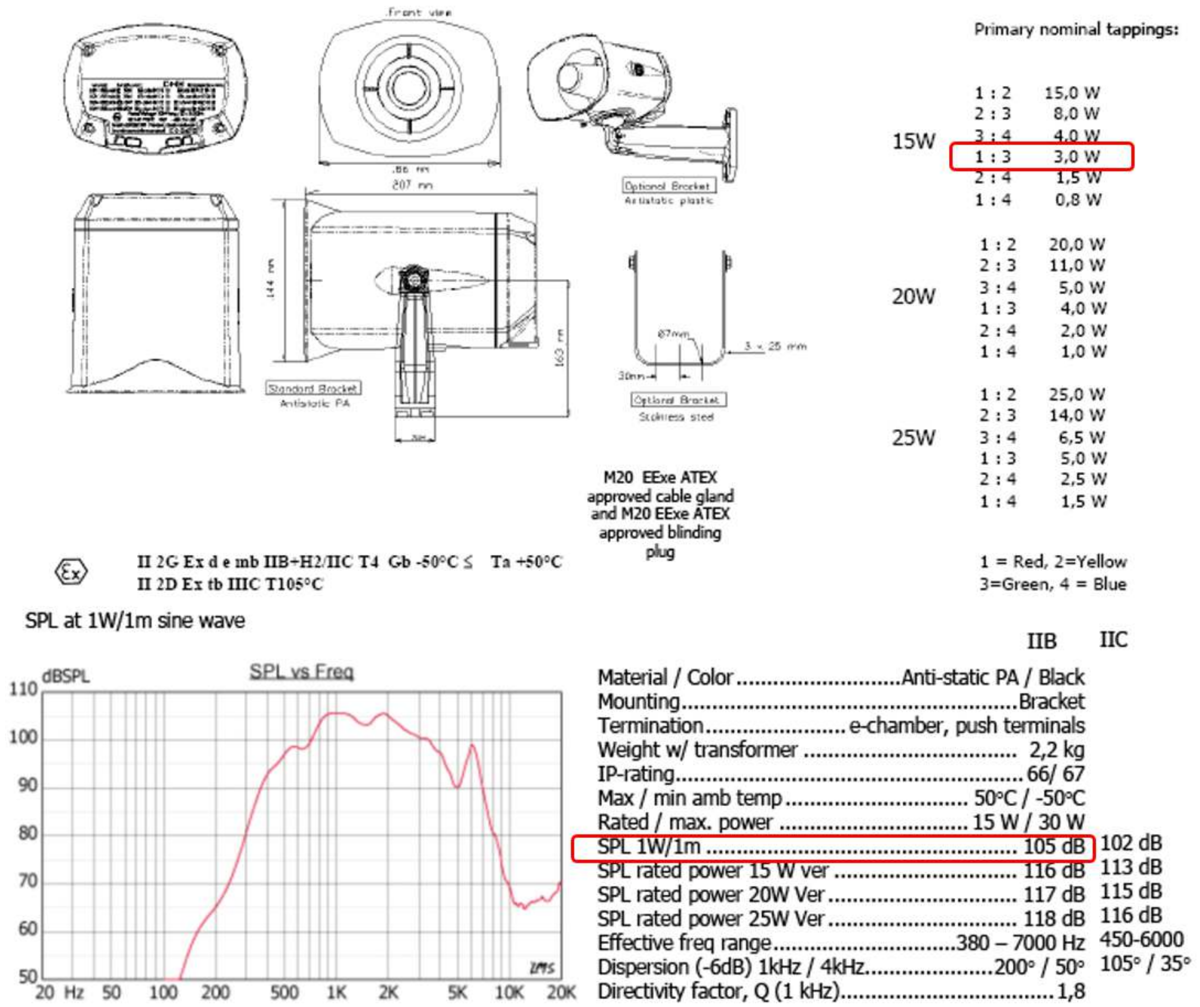

Figure 2 – Example of data sheet for a loudspeaker. This comes in three versions (15W, 20W, 25W) each with six possible tapings. Marked in red: Sensitivity 105 dB SPL 1W/1m and tapping 3W.

An example of a loudspeaker data sheet is shown in Figure 2. The possible tapings for the 15W version correspond to different settings of the overall gain in ODEON as shown in Table 4. This loudspeaker will be applied later in example 3.

Table 4 – Example of the conversion from tapping in W to overall gain in dB.

Tapping W	Overall gain dB
15,0	11,8
8,0	9,0
4,0	6,0
3,0	4,8
1,5	1,8
0,8	-1,0

Receivers

The receiver positions should be placed according to the use of the room. The typical height above the floor is 1,2 m for a seated adult and somewhat lower to represent children. For schools the BB93 [7] specifies the receiver height of 0,8 m for nursery school , 1,0 m for primary school, and 1,2 m for secondary school.

Background noise

The background noise is an important parameter in the STI calculation, especially when the background noise is high. It is specified as background noise SPL in octave bands.

Table 5 – Examples of sound power levels to be used to calculate the overall noise level in an open-plan space for teaching [7].

Frequency, Hz	63	125	250	500	1000	2000	4000	8000
Sound power levels, dB								
Open-plan space - general working (per 15 students)	-	62	62	62	62	57	52	47
Dining space (per 60 students)	-	61	65	69	69	61	51	40
Speech at normal level (per person)	-	60	66	69	62	57	53	50
Quiet student being addressed by the teacher (per student)	-	30	32	32	30	28	26	20

In some cases the background noise must be calculated prior to the STI calculation. The ODEON model of the room can be used modelling the noise sources and their emitted sound power levels. The most common sources of background noise are heating and ventilation equipment or noisy machinery.

In the case of open-plan schools the BB93 guideline [7] provides some examples of typical noise sources due to human activities, see Table 5.

Results

The STI can take values between 0 and 1. In order to evaluate the results, the five step rating shown in Table 6 is often preferred. So, STI > 0.60 is good and is often aimed at in an auditorium. However, the JND (just noticeable difference) for STI is found to be 0.03 by Bradley et al. [8]. So, in the latest edition of IEC 60268-16 [1, Annex F] a scale with 12 qualification bands labelled from A+ to U is suggested, see Table 7. This scale has steps of 0.04 i.e. approximately one JND.

Table 6 – Intelligibility rating and relations to STI and phonetically balanced (PB) word score [2].

Intelligibility rating	STI	Meaningful PB-word score (%)
Excellent	> 0,75	> 98
Good	0,60 to 0,75	93 to 98
Fair	0,45 to 0,60	80 to 93
Poor	0,30 to 0,45	60 to 80
Bad	< 0,30	< 60

Table 7 – Nominal qualification bands for STI [1].

Qualification band	STI
A+	> 0,76
A	0,72 to 0,76
B	0,68 to 0,72
C	0,64 to 0,68
D	0,60 to 0,64
E	0,56 to 0,60
F	0,52 to 0,56
G	0,48 to 0,52
H	0,44 to 0,48
I	0,40 to 0,44
J	0,36 to 0,40
U	< 0,36

Excellent intelligibility rating (STI > 0.75) is unusual in room acoustics; in large and relatively noisy public rooms the intelligibility rating *fair* (0.45 – 0.60) is a realistic goal for a PA system. Examples of recommended minimal performance ratings are shown in Table 8. For VA systems in very noisy working environments like off-shore installations, the rating *poor* (STI > 0.30) is a realistic goal.

Table 8 – Recommended minimal performance ratings for intelligibility and vocal effort after ISO 9921 [2].

Application	Minimum intelligibility rating	Maximum vocal effort
Person-to-person communication (prolonged normal communication)	Good	Normal
Person-to-person communication (critical)	Fair	Loud
Public address in public areas	Fair	Normal
Alert and warning situations (correct understanding of critical words)	Fair	Loud
Alert and warning situations (correct understanding of simple sentences)	Poor	Loud

Example 1 – Human speaker in auditorium

The room used in this example is an auditorium, which is known to be difficult acoustically. It seats 200 people, the volume is approximately 1160 m^3 and the mid frequency reverberation time is 1.8 s.

First the background noise is set in the Room setup menu, see Figure 3. A spectrum dominated by low frequencies is chosen, and the A-weighted SPL is 30 dB(A).

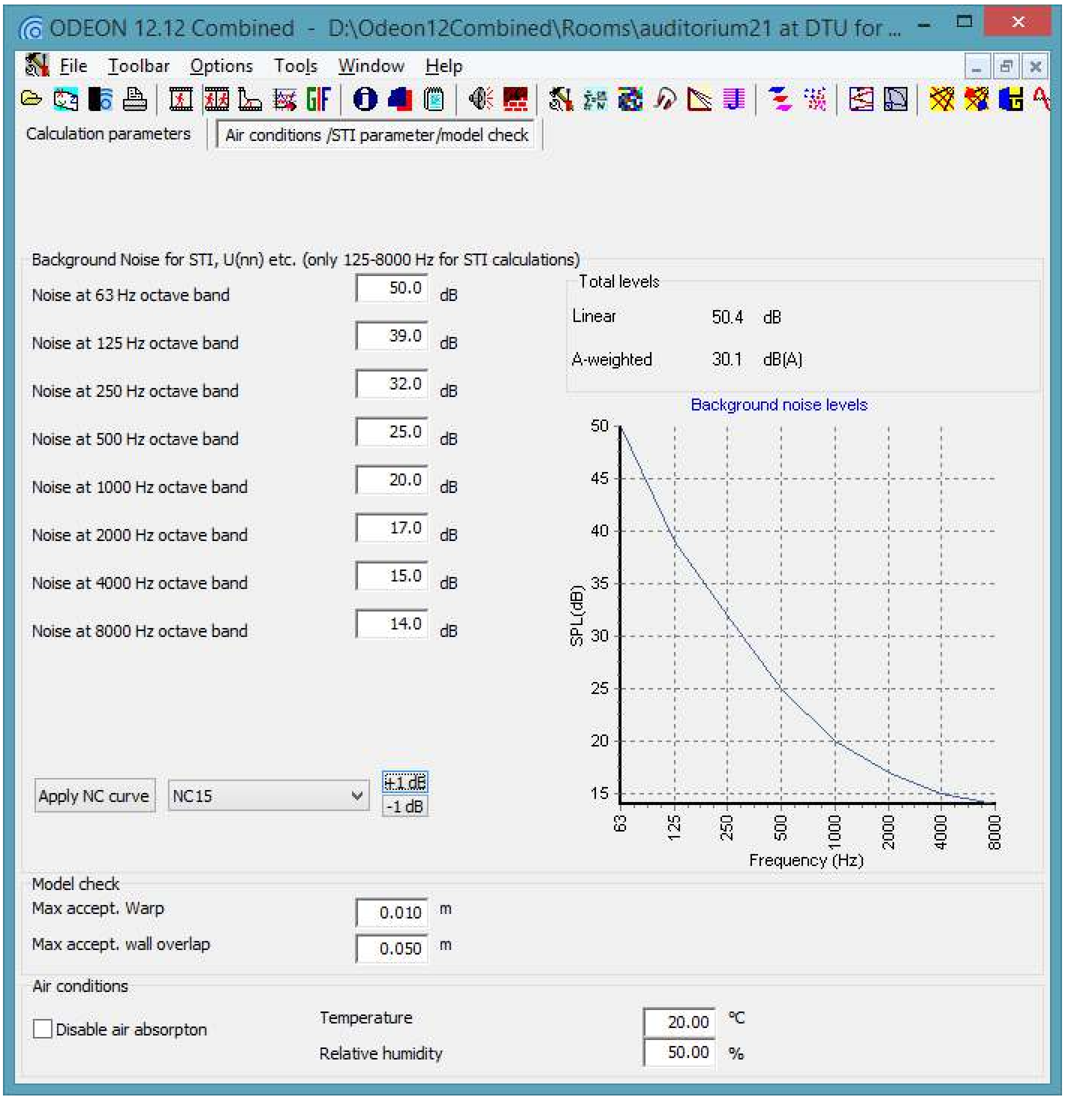

Figure 3 – Setting the background noise in the Room setup in ODEON. In this example a low frequency noise spectrum with 35 dB(A) is assumed.

Next, the sound source is defined. For this example is selected a human speaker with spectrum and vocal effort representing a raised voice, see Figure 4. The SPL on axis at 10 m is 46.5 dB(A), which corresponds to 66.5 dB(A) in a distance of 1 m.

The calculations are made using the grid response. This allows both a visual overview of results, see Figure 5, and at the same time useful statistical data can be derived, see Figure 6.

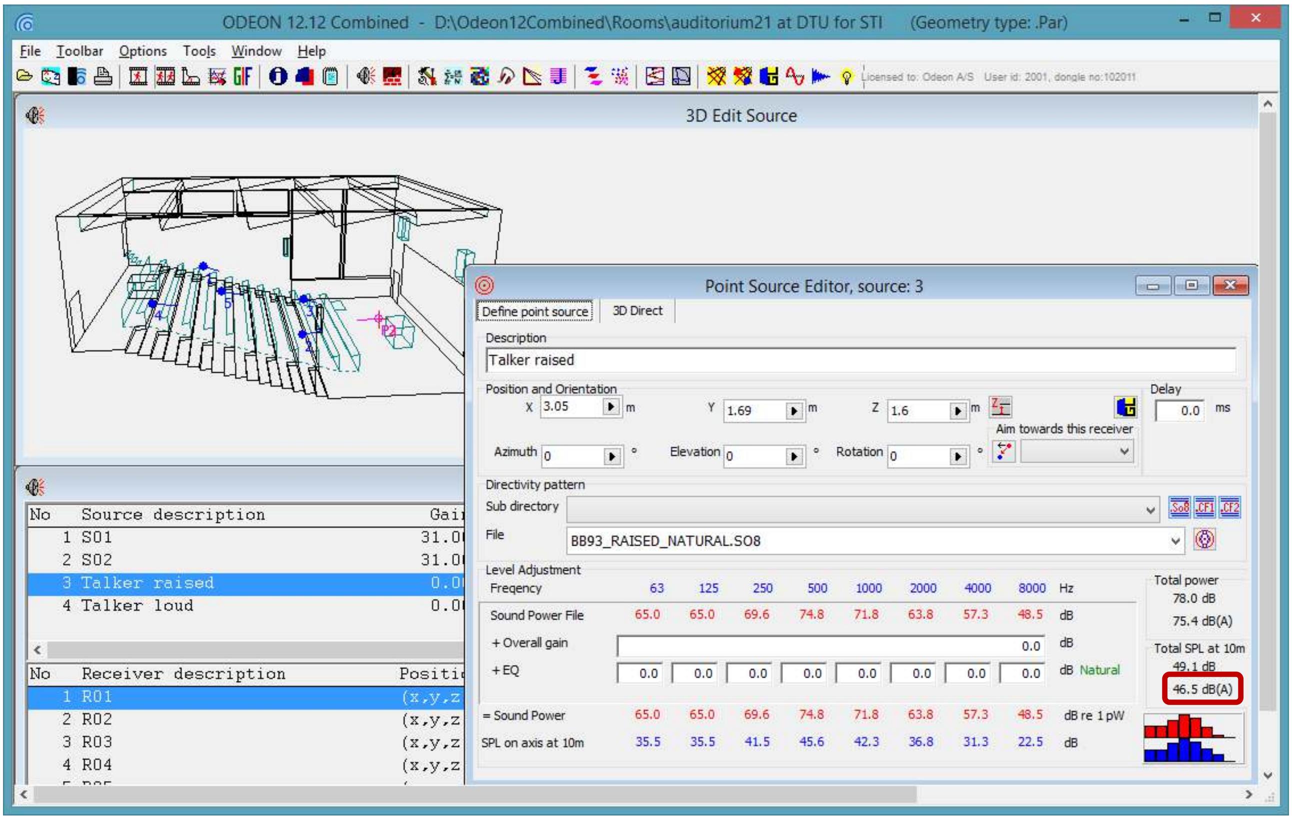

Figure 4 – Defining the source as a human speaker with vocal effort raised. Total SPL at 10 m is 46.5 dB(A).

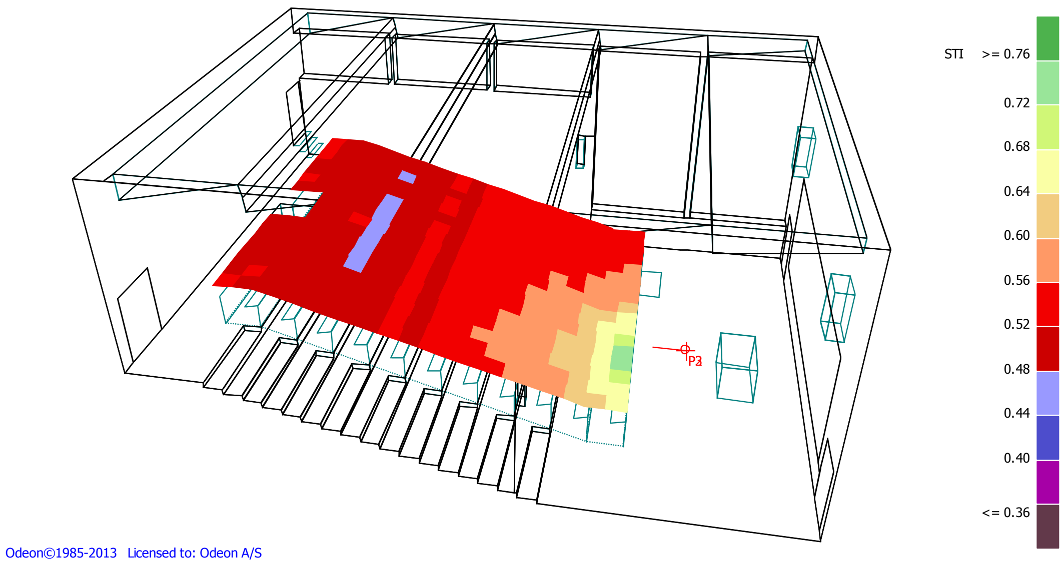

Figure 5 – A grid map showing the STI in the auditorium. The source is a human speaker with vocal effort raised. The results are displayed with the 12-step scale defined in Table 7.

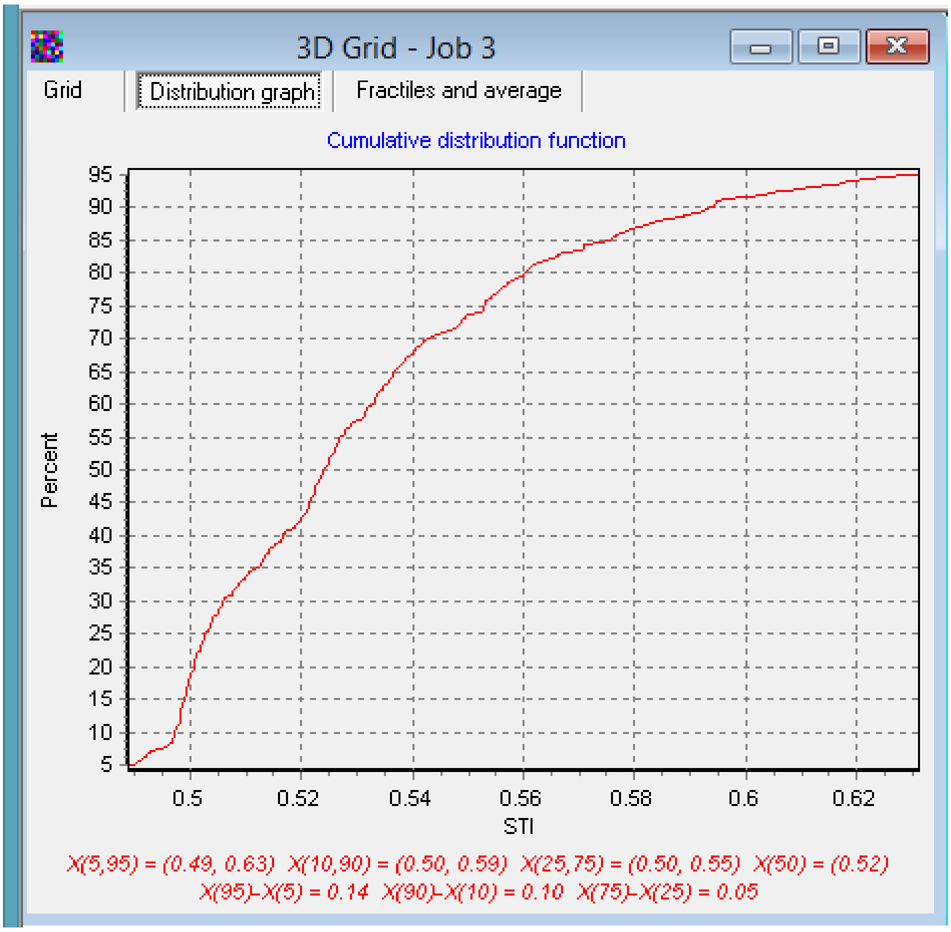

Figure 6 – Statistical data related to the grid map calculation of STI shown in Figure 5. The 5 percentile is 0.49, i.e. STI > 0.49 in 95 % of the grid area. Similarly, the median-value (the 50 percentile) is 0.52.

Example 2 – Sound reinforcement in auditorium

The auditorium from example 1 is used again for this example of a reinforcement loudspeaker system. The background noise and the receiver grid are the same as in the previous example.

Two loudspeakers are used as sound sources. The positions are shown in Figure 7. Figure 8 shows the ODEON source editor menu for one of the loudspeakers. Full access to the loudspeaker data as in Figure 1 is available by clicking the CLF icon. In accordance with the measurement method in IEC 60268-16 [1] the male speech spectrum is applied to the equalizer setting. Thus, this calculation should only be applied for the STI calculation; for other purposes, and especially for auralisation, the speech spectrum should not be set in the equalizer.

The result of the calculation is shown in Figure 9, where the speaking person is omitted.

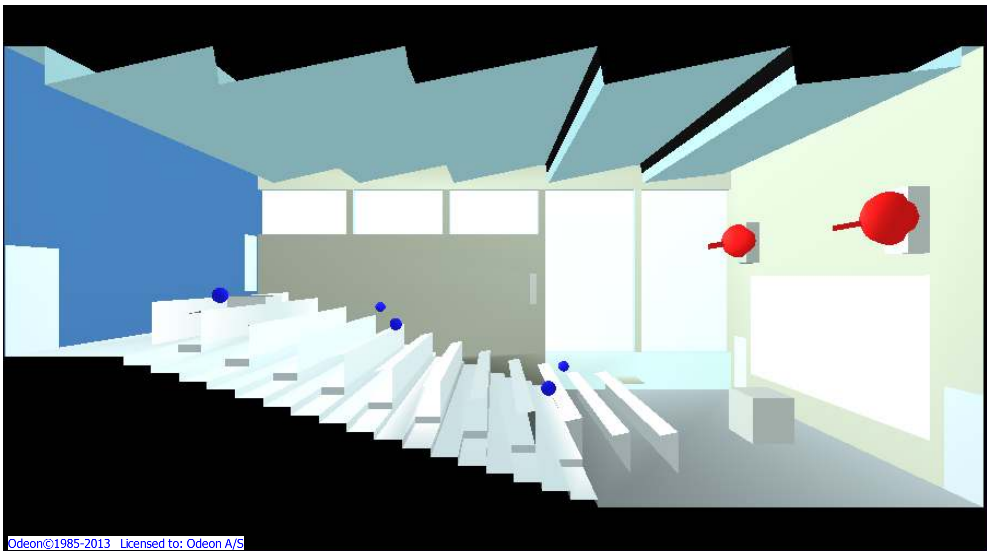

Figure 7 – A view into the model of the auditorium with the sound sources (red) and five receivers (blue).

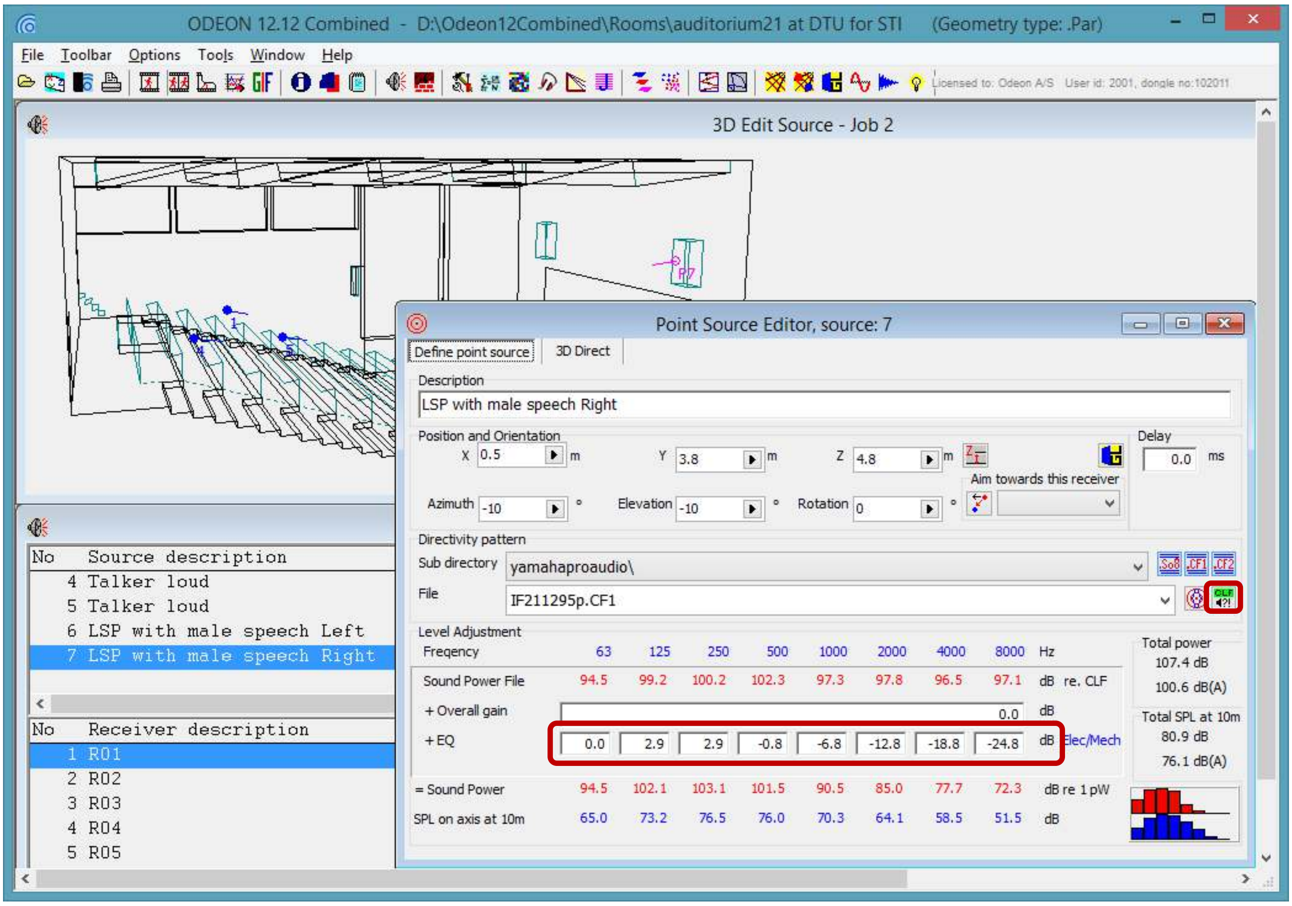

Figure 8 – Defining one of the loudspeakers in the source editor. The source file is the same as in Figure 1, and the detailed information in the CLF1 file is available by clicking the marked CLF icon. Note that the equalizer is set in accordance with IEC 60268-16 [1] with the male speech spectrum from Table 3.

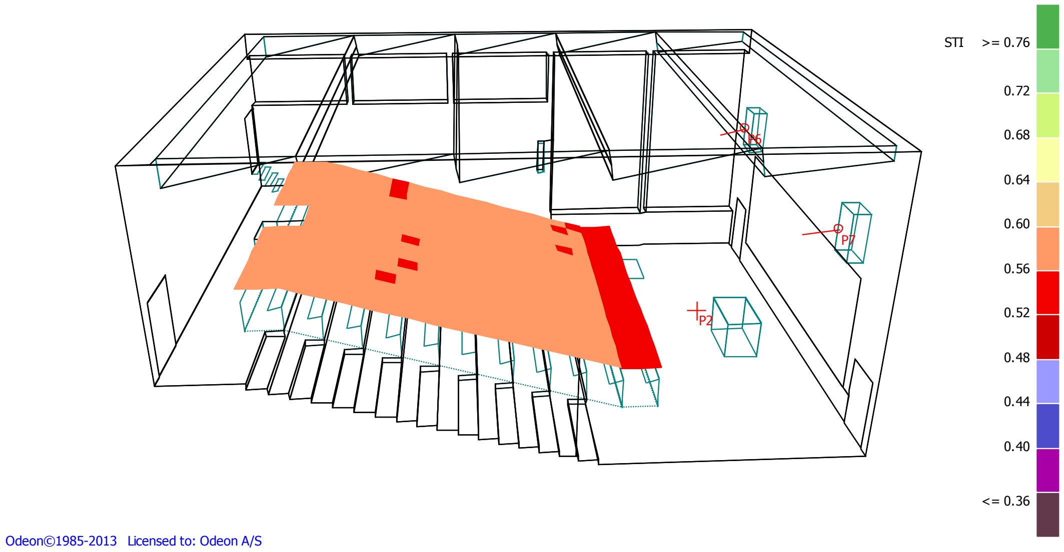

Figure 9 – A grid map showing the STI in the auditorium. The sources are the loudspeakers marked as P6 and P7. From the statistical data related to this grid map, the median-value (the 50 percentile) is 0.58.

Example 3 – PA/VA system

This is an example of calculations with a PA voice alarm system for an off-shore installation.

The background noise is assumed to be 85 dB(A) with a pink noise spectrum, i.e. the same SPL in all octave bands.

The loudspeaker chosen for this case is the same example as in Figure 2. In Figure 10 is seen how one of the loudspeakers is defined in the ODEON source editor. The tapping is chosen to 3W and the gain is set accordingly. Note that for these calculations the equalizer is not used, i.e. a pink noise spectrum is assumed instead of the male speech spectrum. The reason is that this kind of loudspeaker has a very limited frequency range, and sound is mainly emitted in the three octave bands 500 – 2000 Hz.

For security reasons there are two independent loudspeaker installations, and calculations must be made for each of them and for both together. However, only one of them is active in this example.

With the selected position and tapping of the loudspeakers the STI shall be at least 0.35 in the area where people can have access, but at the same time a 1 kHz alarm tone shall produce a SPL that does not exceed 115 dB. The number of loudspeakers, their position and tapping must be selected in order to fulfil these requirements as closely as possible. Both SPL and STI results are shown in Figure 11. The statistical data from the grid response calculations is particularly useful for this kind of calculation.

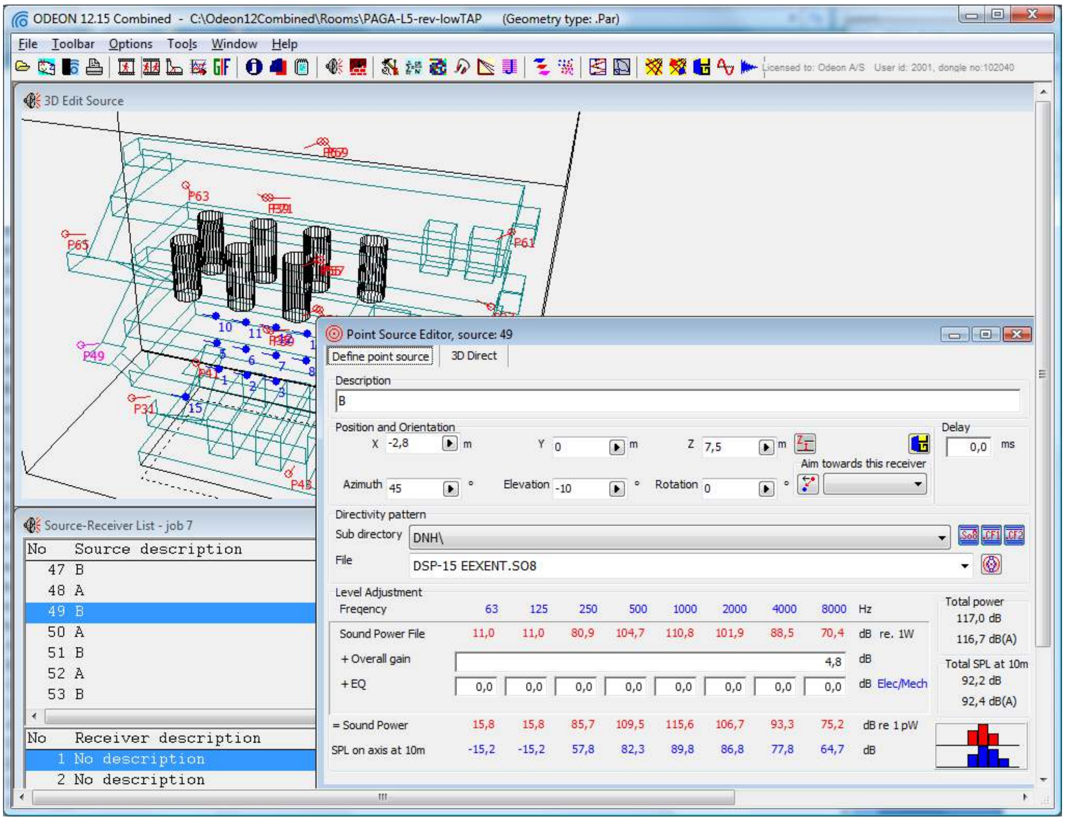

Figure 10 – How to define the loudspeaker from Figure 2 as a sound source in ODEON. The overall gain is set to 4.8 dB corresponding to a tapping of 3W, see Table 4.

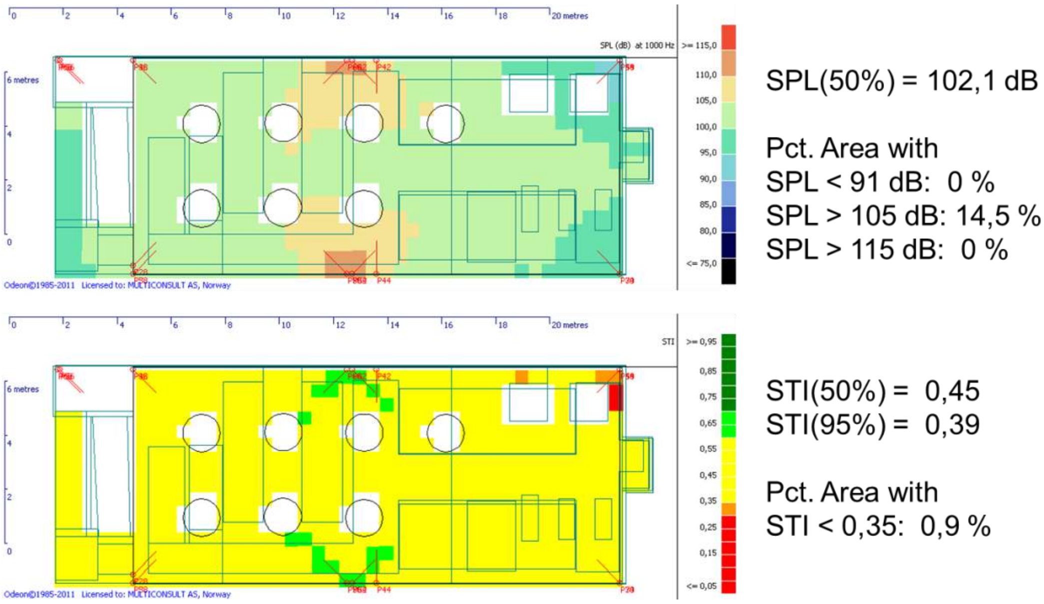

Figure 11 – Calculation results for a PA voice alarm system presented in grid response plots supplemented with results from the cumulative distribution function. Upper part: the SPL at 1 kHz. Lower part: the STI. NB: In this example, only half of the loudspeakers are active.

References

1. IEC 60268-16:2011 *Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index*. International Electrotechnical Commission, Geneva, Switzerland (2009).
2. ISO 9921:2003 *Ergonomics – Assessment of speech communication.* International Organization for Standardization, Geneva, Switzerland (2003).
3. M.R. Schroeder, *Modulation Transfer Functions: Definition and Measurement*. Acustica **49**, 179-182 (1981).
4. ANSI 3.5-1997. *American National Standard – Methods for Calculation of the Speech Intelligibility Index*, (1997).
5. W.T. Chu, A.C.C. Warnock, *Detailed Directivity of Sound Fields Around Human Talkers*, IRC-RR 104, National Research Council, Canada (2002).
6. J.H. Rindel, C.L. Christensen, A.C. Gade, *Dynamic sound source for simulating the Lombard effect in room acoustic modeling software*. Proceedings of Internoise 2012, New York, USA (2012).
7. BB93, *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. Available from: www.teachernet.gov.uk/acoustics.
8. J.S. Bradley, R. Reich, S.G. Norcross, *A just noticeable difference in C50 for speech*, Appl. Acoust. **58**, 99–108 (1999).