ODEON APPLICATION NOTE – Auralisation and how to calibrate the sound level for presentations

JHR, December 2015

Scope

This is a guide how to calibrate the sound level of an auralisation derived from a room acoustic simulation. The correct level of presentation is important for all kinds of auralisation in order to be as realistic as possible.

The method is approximate, aiming at the A-weighted sound pressure level to match that of the calculation result of the same simulation. A pink noise signal is applied for the calibration. The method is straight forward for presentations through loudspeakers, but not easy to adapt to presentations through headphones. This application note refers to ODEON version 13, auditorium or combined editions.

Method

The principle is to replace the sound signal used for the auralisation with a pink noise calibration signal. When played in the setup for listening to the auralisation, the level is adjusted in such a way that the pink noise signal creates a certain pre-specified sound pressure level.

In order to ensure a proper calibration it is necessary to know the sound power level of the sound source used as input for the auralisation. This can be difficult in some cases and may be the major source of uncertainty in the calibration. Next step is to look at the room simulation that produces the auralisation, and to choose a source-receiver combination to be used for the calibration. The relation between the sound power of the source and the sound pressure level in the receiver position is a result of the room acoustic simulation. The third step is to replace the anechoic recording to be used for the auralisation with a pink noise signal that is more suited for the calibration, and to note the level difference introduced by this shift of input signal. Finally, it may be advantageous to apply a correction to the calibration signal in order to get a calibration level that is convenient for the purpose, e.g. 80 dB

The anechoic recordings used for auralisation can be of various types, including:

- speech
- noise
- music (single recording)
- orchestra (multiple recordings of all instruments)

Requirement: The anechoic recordings should be as realistic as possible for the actual purpose. Especially speech recordings should be made with the correct level of vocal effort, because the spectrum and other characteristics of the speech changes with the vocal effort. Unfortunately, the currently available anechoic speech examples are at normal vocal effort.

A large number of anechoic recordings of single musical instruments are available, and a few orchestra examples with all instruments recorded individually are also available.

Definitions

$L_{W,A,1}$
A-weighted sound power level of the source used in calculations in the simulation.

$L_{p,A,1}$
A-weighted sound pressure level at the receiver position as calculated in the simulation.

$L_{W,A,2}$
A-weighted sound power level of the real source in the anechoic recording used for the auralisation.

$L_{p,A,2}$
A-weighted sound pressure level in the receiver position as it should be when listening to the auralisation.

$L_{p,A,2,cal}$
A-weighted sound pressure level of the calibration signal.

ΔL_{cal}
Difference in "*Max. out*" when the convolution is made with or without the "*calib.*" setting.

Auralisation method

Before any calculations are made it is required to choose how to do the auralisation, either by headphones (binaural) or by loudspeakers (2D surround sound). As the calibration method described here is meant for loudspeaker presentation, this should be selected in the auralisation setup, see Figure 1. The overall recording level may be changed later, if needed.

It is also possible to choose both binaural and surround sound, although the calculation time will then be somewhat longer.

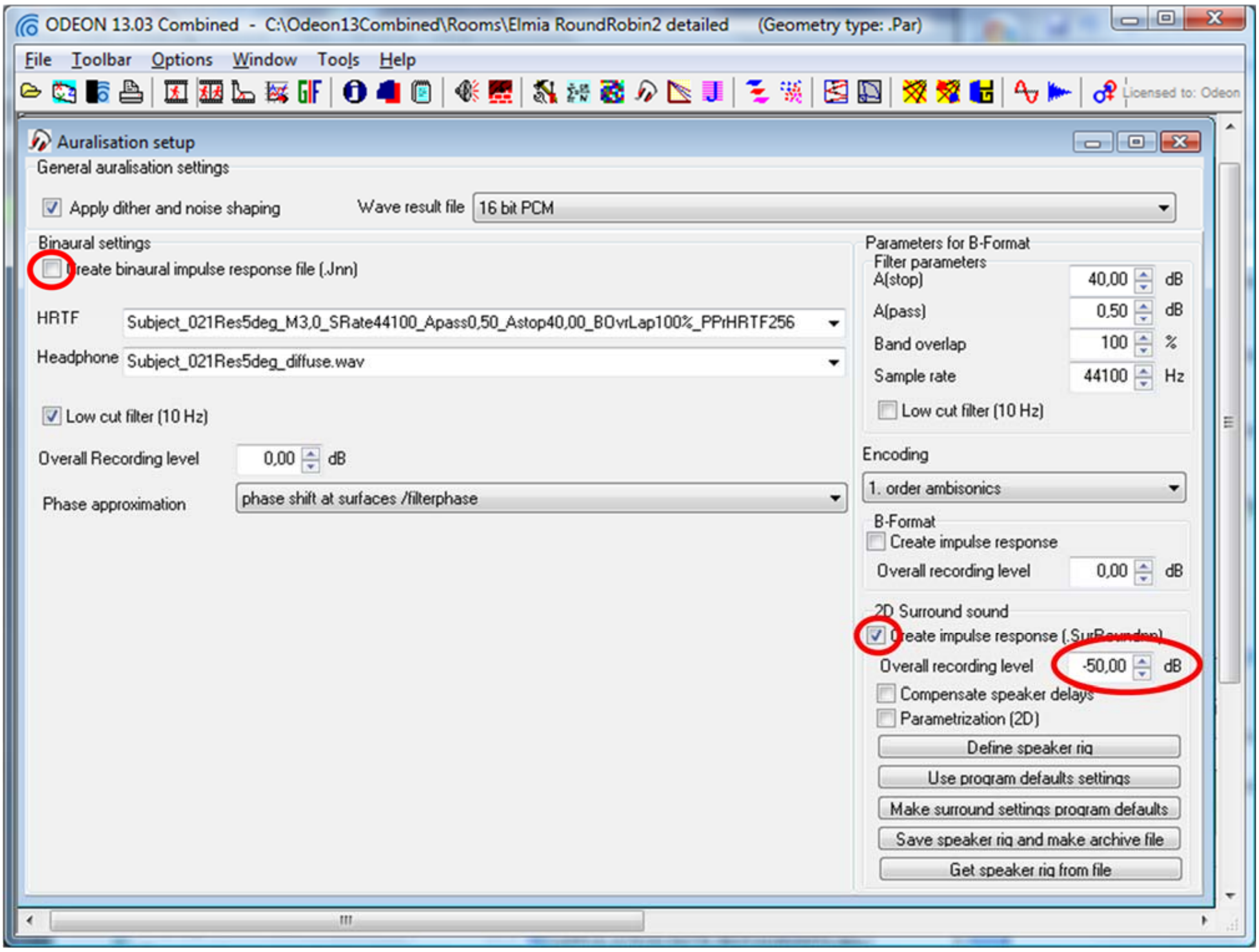


Figure 1. The auralisation setup. Here the loudspeaker presentation is selected, while the binaural presentation through headphones is disabled (upper left corner).

A person speaking in an auditorium

The following section describes the calibration of a single speech signal in a room.

Speech levels can be taken either from ANSI 3.5 or from ISO 9921. ANSI 3.5 has four levels of vocal effort from normal to shouted, see Table 1. The original data are given as SPL at a distance of 1 m in front of the mouth; for applications to room acoustic simulations the source data should be as sound power levels. This conversion to sound power levels is further described in [3].

Table 1. Speech spectra given as sound power levels in octave bands for different vocal effort. Based on data from ANSI 3.5 [1]. The sound power levels at 63 and 125 Hz are from Rindel et al. 2012 [3].

	63	125	250	500	1000	2000	4000	8000	dB(A)
Normal	45.0	55.0	65.3	69.0	63.0	55.8	49.8	44.5	**68.4**
Raised	48.0	60.0	69.5	74.9	71.9	63.8	57.3	48.4	**75.5**
Loud	52.0	64.0	72.1	79.6	80.2	72.9	65.9	54.8	**82.6**
Shouted	52.0	65.0	73.1	84.0	89.3	82.4	74.9	64.1	**90.9**

ISO 9921 [2] defines six different levels of vocal effort from *relaxed* to *very loud*, and the step size between levels is 6 dB, see Table 2. The data are given as A-weighted sound pressure levels in a distance of 1 m in front of the mouth. Approximately 9 dB should be added to convert these data into A-weighted sound power levels.

Table 2. Vocal effort of a male speaker and related SPL(A) at 1 m in front of the mouth. After ISO 9921[2].

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

A sound source is selected for the calculation. In the example in Figure 2 the source “BB93 RAISED NATURAL.SO8” is selected. It has the spectrum and the sound power level corresponding to *raised* vocal effort, see Table 1 (small and unimportant differences at the low frequencies). The source also has the directivity of a talking human person. The source is positioned and oriented in the room model and used for the calculation of a point response at a chosen receiver position; see the job setup in Figure 3.

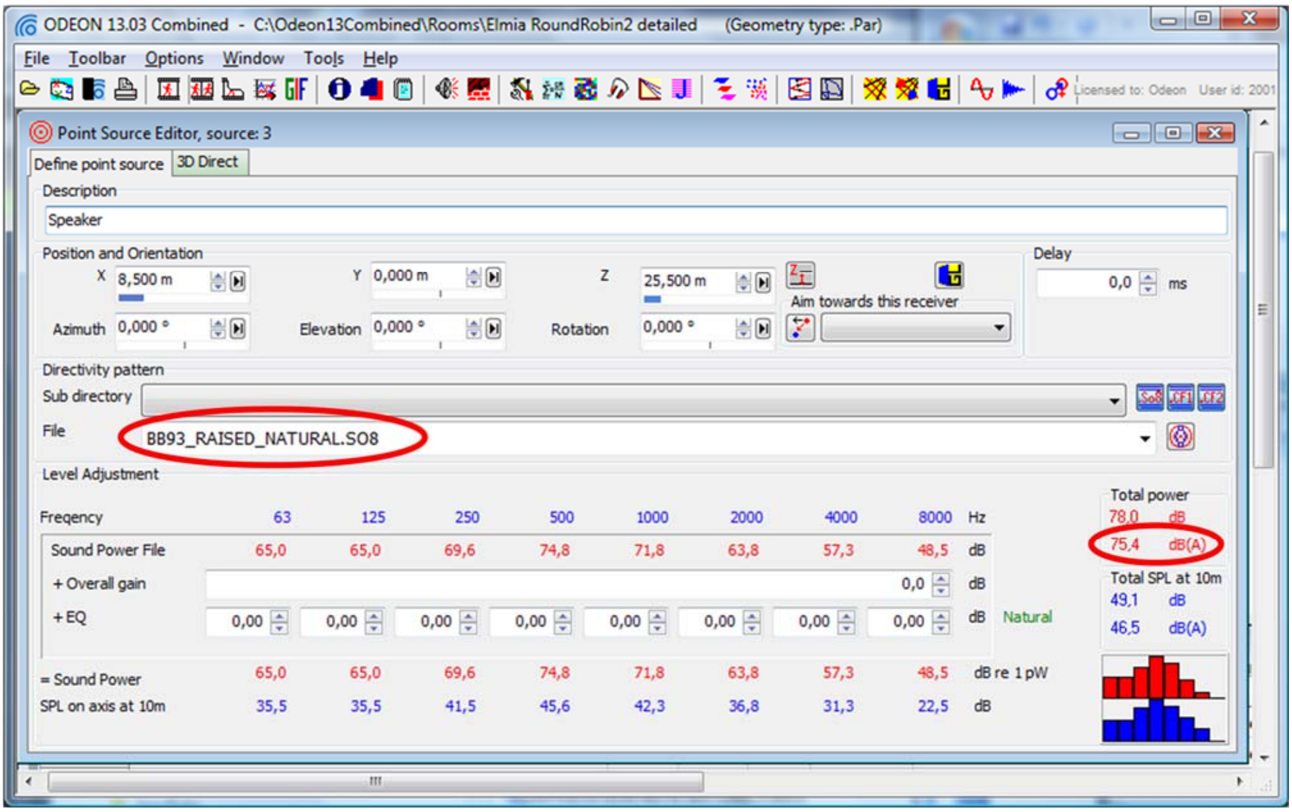


Figure 2. The point source is selected (speech spectrum and directivity) *raised* vocal effort. The sound power level is 75.4 dB(A). The SPL at 10 m is 46.5 dB(A), which means that $L_{S,A,1m}$ = 66.5 dB(A).

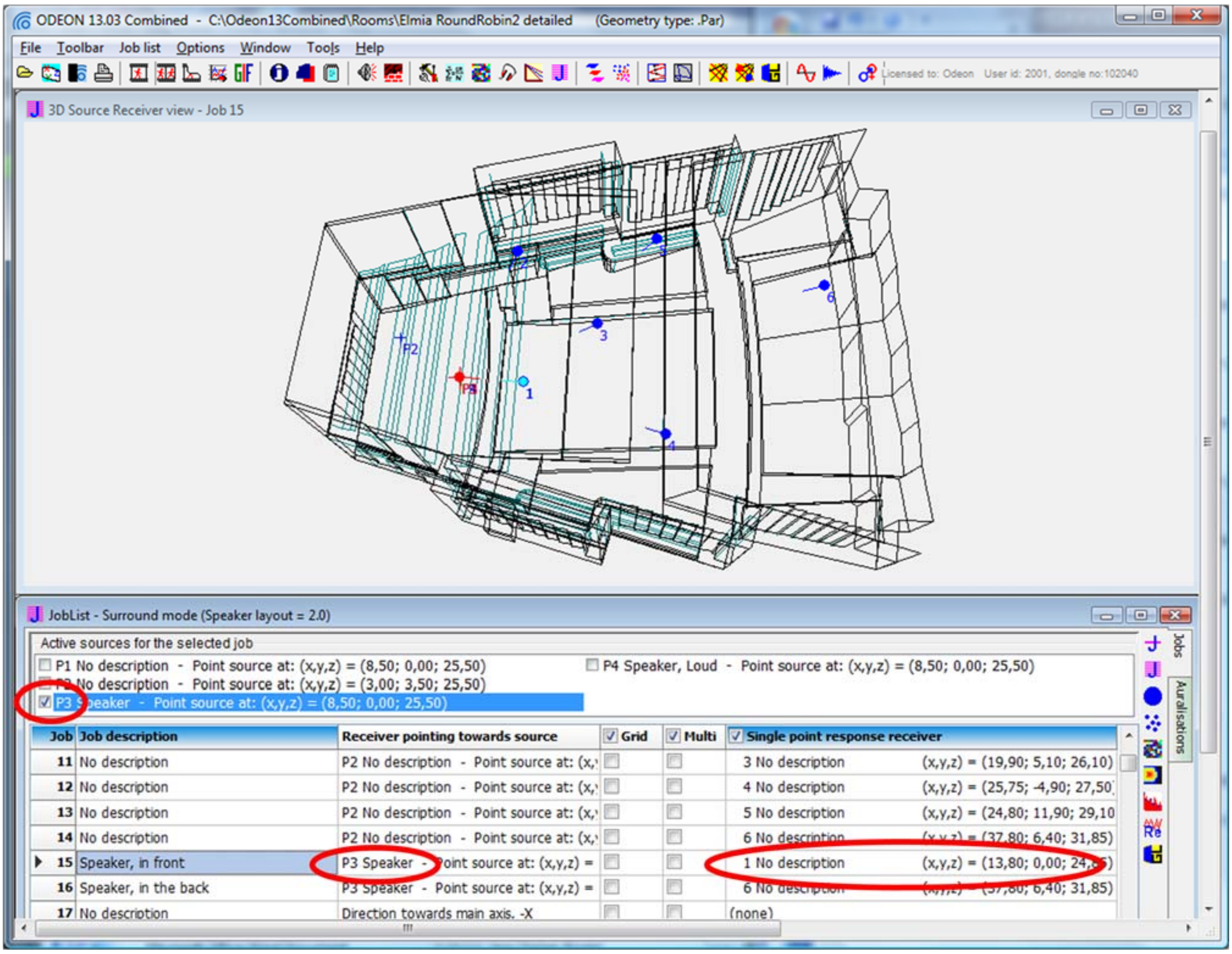
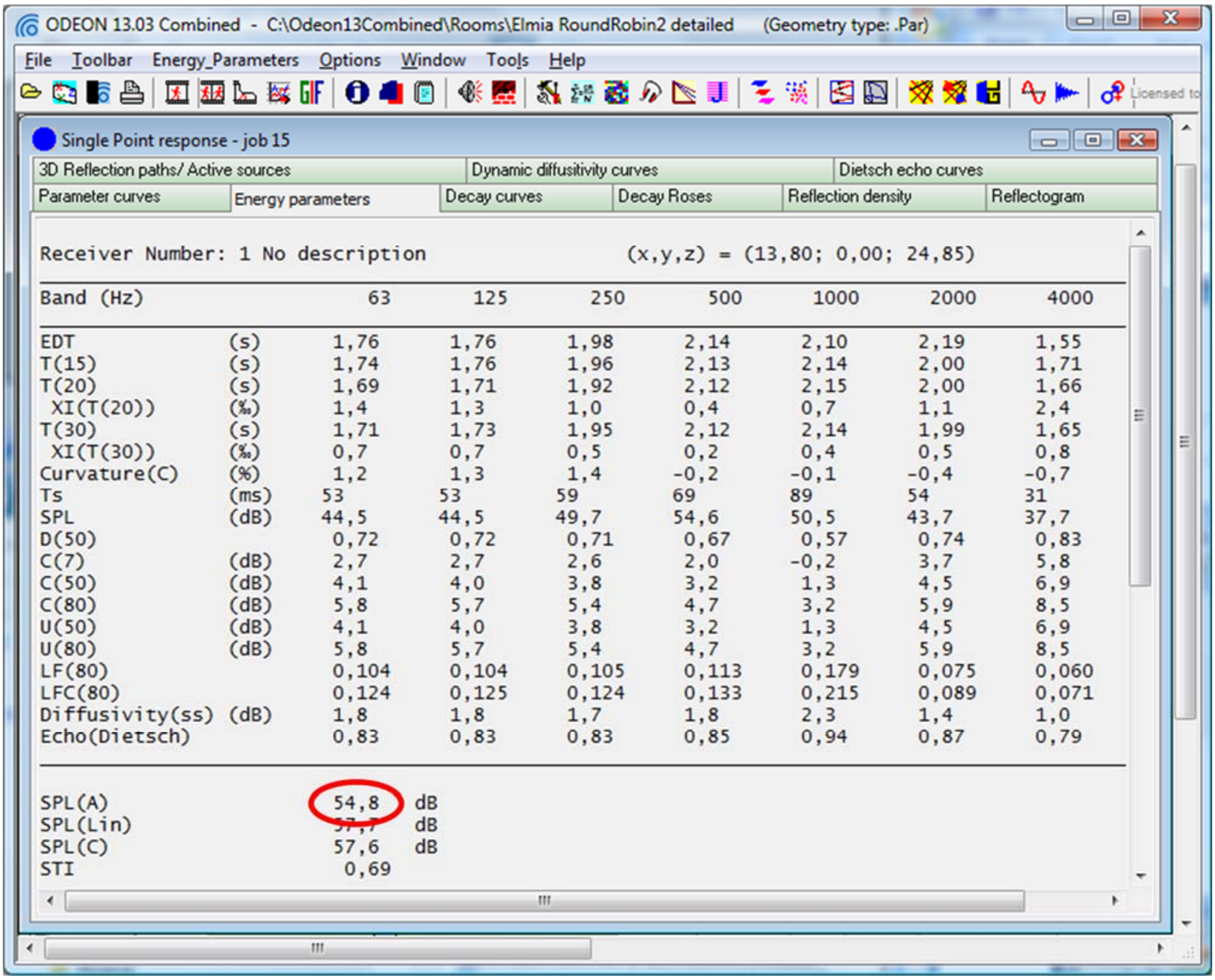


Figure 3. A job (15) is defined with source P3 active and receiver no. 1 pointing towards P3. Another job (16) is prepared with receiver no. 6.

A table with the result of the calculated point response is seen in Figure 4. The A-weighted sound pressure level is 54.8 dB, so this is the level we shall aim at in the presentation of an auralisation listening to speech, if the vocal effort is assume to be *raised*.



Receiver Number: 1 No description (x,y,z) = (13,80; 0,00; 24,85)

Band (Hz)		63	125	250	500	1000	2000	4000
EDT	(s)	1,76	1,76	1,98	2,14	2,10	2,19	1,55
T(15)	(s)	1,74	1,76	1,96	2,13	2,14	2,00	1,71
T(20)	(s)	1,69	1,71	1,92	2,12	2,15	2,00	1,66
XI(T(20))	(‰)	1,4	1,3	1,0	0,4	0,7	1,1	2,4
T(30)	(s)	1,71	1,73	1,95	2,12	2,14	1,99	1,65
XI(T(30))	(‰)	0,7	0,7	0,5	0,2	0,4	0,5	0,8
Curvature(C)	(%)	1,2	1,3	1,4	-0,2	-0,1	-0,4	-0,7
Ts	(ms)	53	53	59	69	89	54	31
SPL	(dB)	44,5	44,5	49,7	54,6	50,5	43,7	37,7
D(50)		0,72	0,72	0,71	0,67	0,57	0,74	0,83
C(7)	(dB)	2,7	2,7	2,6	2,0	-0,2	3,7	5,8
C(50)	(dB)	4,1	4,0	3,8	3,2	1,3	4,5	6,9
C(80)	(dB)	5,8	5,7	5,4	4,7	3,2	5,9	8,5
U(50)	(dB)	4,1	4,0	3,8	3,2	1,3	4,5	6,9
U(80)	(dB)	5,8	5,7	5,4	4,7	3,2	5,9	8,5
LF(80)		0,104	0,104	0,105	0,113	0,179	0,075	0,060
LFC(80)		0,124	0,125	0,124	0,133	0,215	0,089	0,071
Diffusivity(ss)	(dB)	1,8	1,8	1,7	1,8	2,3	1,4	1,0
Echo(Dietsch)		0,83	0,83	0,83	0,85	0,94	0,87	0,79

SPL(A)	54,8	dB
SPL(Lin)	57,7	dB
SPL(C)	57,6	dB
STI	0,69	

Figure 4. Point response result for job 15.

In this example the sound source used in the simulation has the correct sound power, i.e. $L_{W,A,1} = L_{W,A,2} =$ 75.4 dB.

The sound pressure level in the receiver position is calculated to be $L_{p,A,1} = L_{p,A,2} = 54.8$ dB ≈ 55 dB, which is also the level we should try to obtain in the calibrated auralisation.

The calibration signal itself should preferably be sufficiently loud to make the calibration easy. Here we choose an A-weighted sound pressure level of the calibration signal $L_{p,A,2,cal} = 80$ dB. This means that the calibration signal must be 80 – 55 = 25 dB above the speech signal (see convolution 1 and 2 in Figure 5). We could simply attenuate the speech signal by 25 dB, but in order to get close to the maximum output of the sound card, the noise signal is adjusted up by 10 dB, while the speech signal is adjusted down by 15 dB.

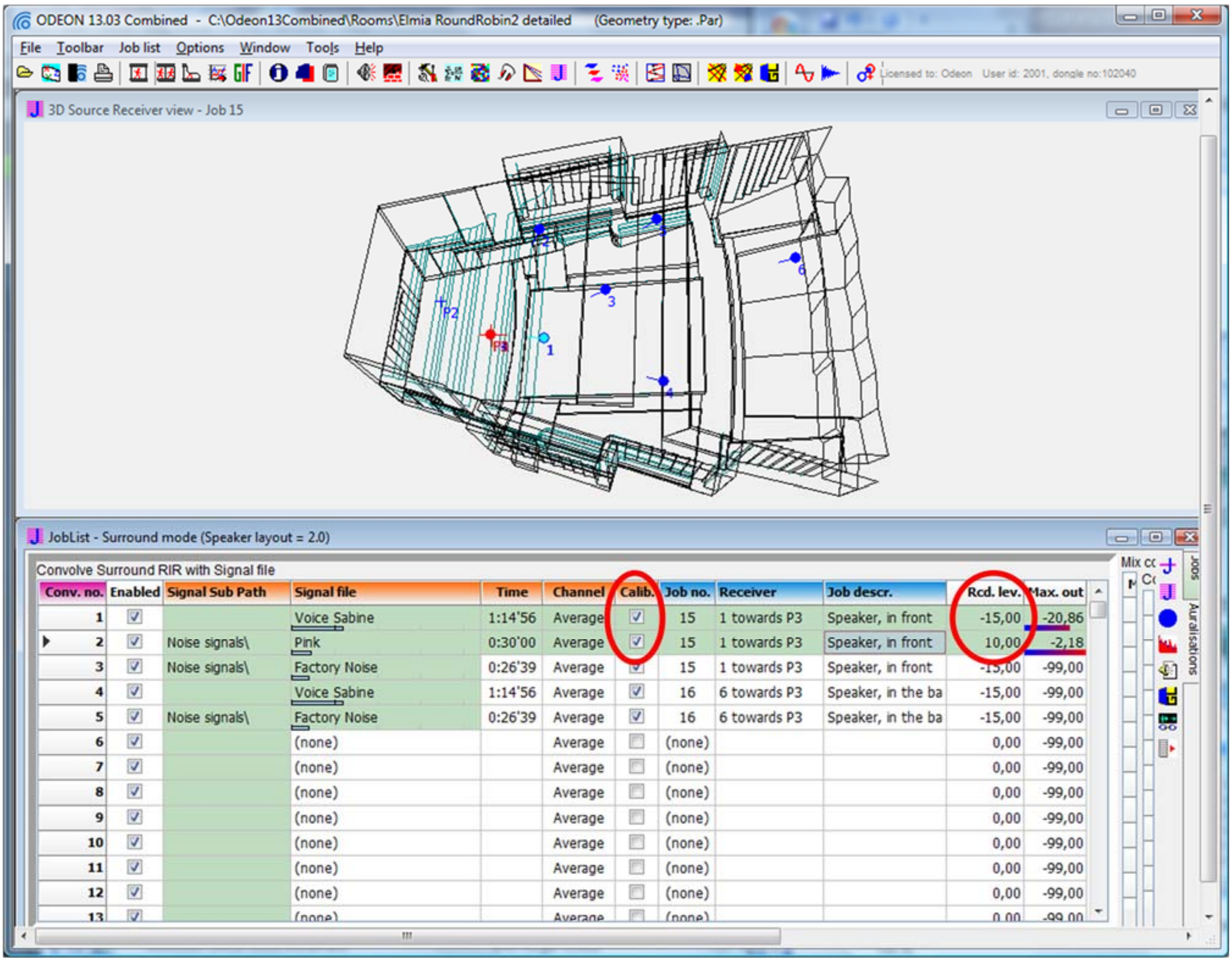


Figure 5. The auralisation page. Convolution 1 and 2 are with speech and noise signal, respectively, and both using the impulse response from job 15. Note, that the Calib. is activated.

When the *calib.* check mark is checked(see Figure 5), ODEON will 'calibrate' the level of the input to have a fixed A-weighted level; so with two different recordings, e.g. voice and noise, their loudness will become comparable (they are adjusted to have the same A-weighted level before they are convolved with the impulse response).

Now the sound system should be turned on. The convolved pink noise (convolution no. 2) is then played. Finally, the volume control should be adjusted in order to get 80 dB(A) at a listening position in the room, where the auralisation is presented. When done, the calibration is finished and the presentations can start.

In this example the presentation is the speech (loud vocal effort) as heard in receiver position 1. The auralisation is played (convolution no. 1) and the average A-weighted should be close to 55 dB(A) as calculated.

Other receiver positions can be chosen and the source can be moved to other positions as well. Any new auralisation will automatically be calibrated as long as the *calib.* is checked and the recording level is kept at -15 dB, as in Figure 5. It is even possible to choose another anechoic recording, and the calibration is still valid.

Mixing convolutions

Next case is the auralisation of a noisy restaurant, by mixing several convolutions together. The listener is supposed to have a conversation with a talking person at a distance of 1.0 m in front of the listener, but the restaurant is crowded with talking people that creates a significant ambient noise. The talking person is modelled as a point source with speaker directivity and spectrum, and the ambient noise is modelled as a surface source, also with speech spectrum. For further details see the ODEON Application Note on *Acoustics in Restaurants* [4].

In Figure 6 is seen the auralisation page. Job 6 is with the surface source and job 7 is with the point source at 1 m distance. The A-weighted sound pressure level at the receiver point is 81.0 dB in job 6 and 68.1 dB in job 7. Thus, there is a difference of 13 dB, which must be compensated for. In order to optimize the use of the dynamic range of the sound card, the talker convolutions (job 7) are increased by 3 dB, and the ambient noise (job 7) are decreased by 10 dB.

Three different sounds are used for the convolutions; crowd noise (PARTY1 - short), male speech (Agora short), and pink noise (Pink), see Figure 6. The *Calib.* is checked in all cases, which means that the relative difference in SPL is controlled by adjusting the level of each convolution.

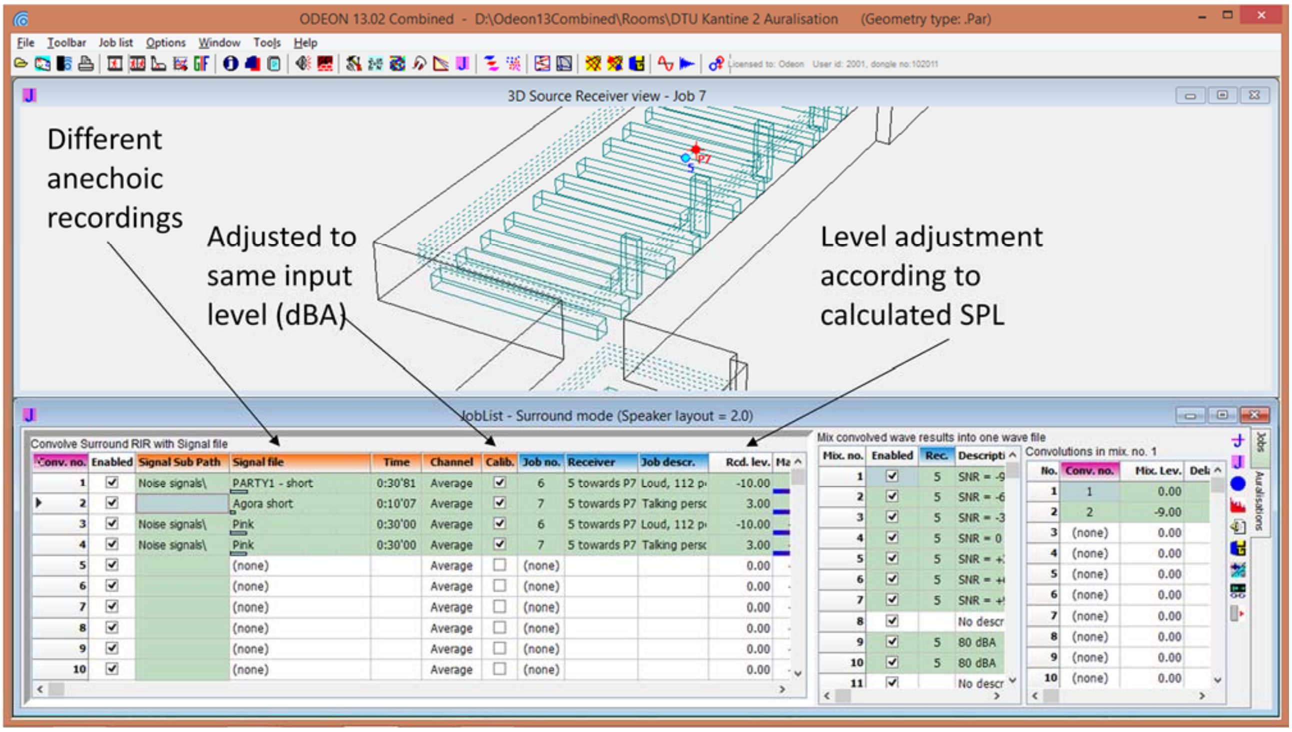


Figure 6. The auralisation page. Convolution 1 is with crowd noise, convolution 2 is with speech and convolutions 3 and 4 are with pink noise signal. Note, that the *Calib.* is activated.

In this case the aim is to auralize different quality levels of verbal communication (see Table 3). The highest level to be presented is 83 dB. Thus this level is assumed to correspond to the maximum output of the convolutions. In the convolution mixer the ambient noise and the speech signal are combined and the *Mix. Lev.* is used to adjust each signal to the actual wanted level below 83 dB.

Table 3. The sound pressure levels of speech and ambient noise and the associated quality of verbal communication.

Quality of verbal communication	SNR dB	$L_{S,A,\,1m}$ dB	L_{NA} dB
Very good			
	9	56	47
Good			
	3	62	59
Satisfactory			
	0	65	65
Sufficient			
	-3	68	71
Insufficient			
	-9	74	83
Very bad			

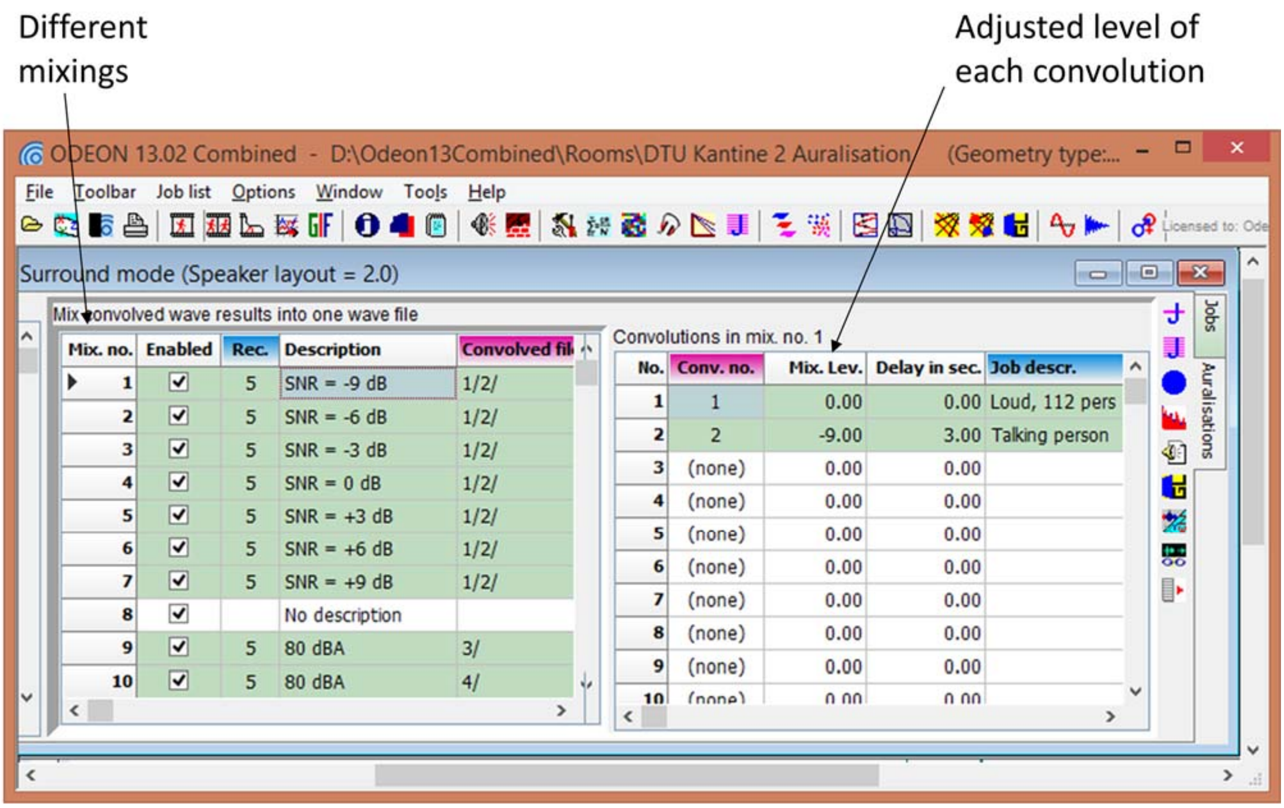


Figure 7. The auralisation mixer. The mix. level of 0 dB corresponds here to a presentation level of 83 dB.

In Figure 7 is shown mix no. 1 having the speech signal adjusted by -9 dB, i.e. to create 74 dB. Similarly is set for other levels in the other mixes. Mix no. 9 and 10 are the pink noise signals in convolution 3 and 4, but with -3 dB adjustment; these signals will then represent at 80 dB.

When auralizing these convolutions the pink noise signal (Mix. no. 10) is played first, and the level of the loudspeakers is adjusted to give 80 dB(A) at the listening position in the room. After that, the other auralisation mixes can be presented with correct levels.

Orchestra

The calibration of the auralisation of a symphony orchestra is very difficult and only possible to some degree. However, a suggestion for an approximate calibration procedure is given in the following. As an example we look at the Brahms 4th symphony, 3rd movement, setup in a model of the Concertgebouw Concert hall. For further details of the setup, see the ODEON Application Note on *Orchestra simulation and auralisation* [5]. The orchestra setup on the stage and some receiver positions are shown in Figure 8.

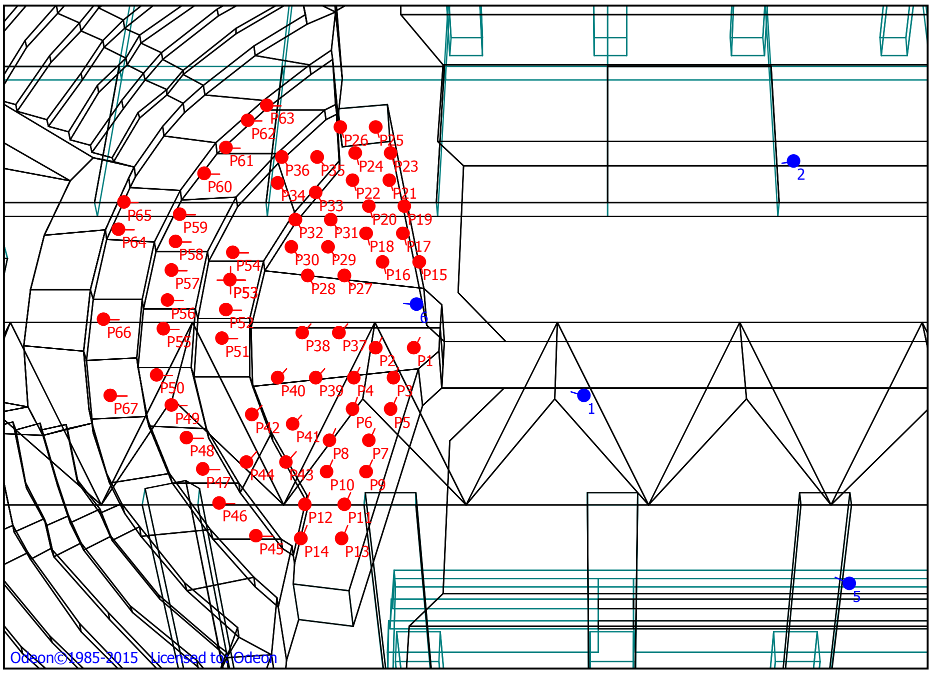


Figure 8. Orchestra setup in the concert hall model. The 1st oboe is source P53. Auralisation is made for receiver no. 2.

The sound level created by an orchestra depends on many factors:

- Kind of instrument
- Number of instruments
- Dynamic level of music being played (*piano, forte,* etc.)
- Acoustic support of the room and listener position (the strength, *G*)

For instance, string instruments are weaker than woodwind instruments, which are again weaker than brass instruments and percussion. Some information about this can be found in [6]. Each instrument has a dynamic range which is typically 20 – 30 dB. As an example, Table 4 shows the sound power levels of an oboe at five different dynamic levels.

Table 4. Average sound power level of the oboe at various dynamic levels. After Meyer [6].

Dynamic level	*pp*	*p*	*mf*	*f*	*ff*
$L_{W,A}$ (dB)	81	85	89	93	97

For the calibration we chose one instrument, only, and the 1st oboe may be a good choice for the purpose. If no details are available about the actual music piece, *forte* level is assumed, i.e. $L_{W,A,2}$ = 93 dB.

However, a more detailed evaluation is possible in this case. If we look into the music score, we notice that the dynamic notation is *fortissimo* (*ff)*, see Figure 9*.* This means plus 4 dB according to Table 4. However, if we then look at the total wav-file of this particular recording in Figure 10, it appears that the oboe plays roughly only half of the time. This means minus 3 dB.

So we end up with the sound power level $L_{W,A,2}$ = 93 + 4 – 3 dB = 94 dB.

Figure 9. Score showing the first 13 bars of the Brahms music example. The dynamic notation is *ff*.

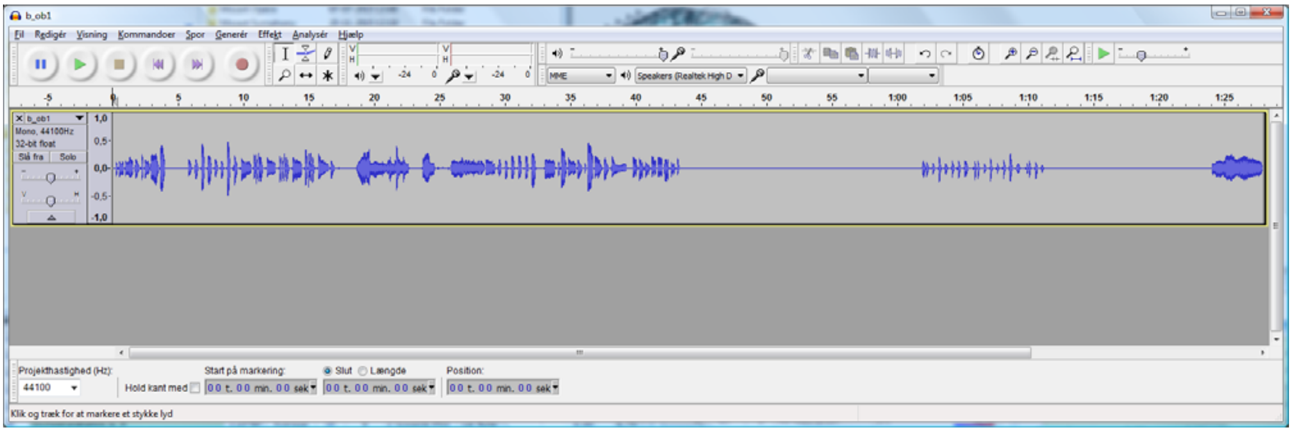


Figure 10. The sound file “b_ob1” (1st oboe in the Brahms symphony) anechoic recording.

The sources representing musical instruments and used for the calculations are all defined with a sound power level $L_{W,A,1}$ = 7.0 dB(A), see Figure 11. The calculation result for this source in receiver position R2 is shown in Figure 12, and the sound pressure level is $L_{p,A,1}$ = -18.3 dB(A).

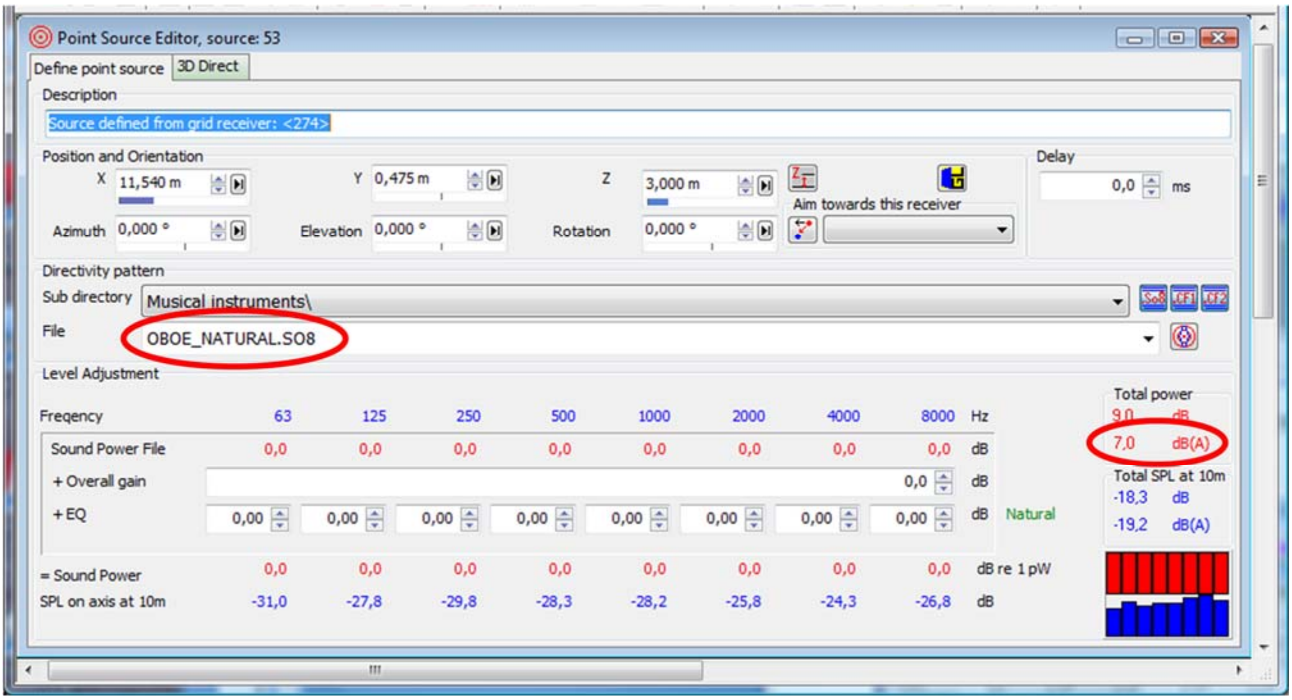
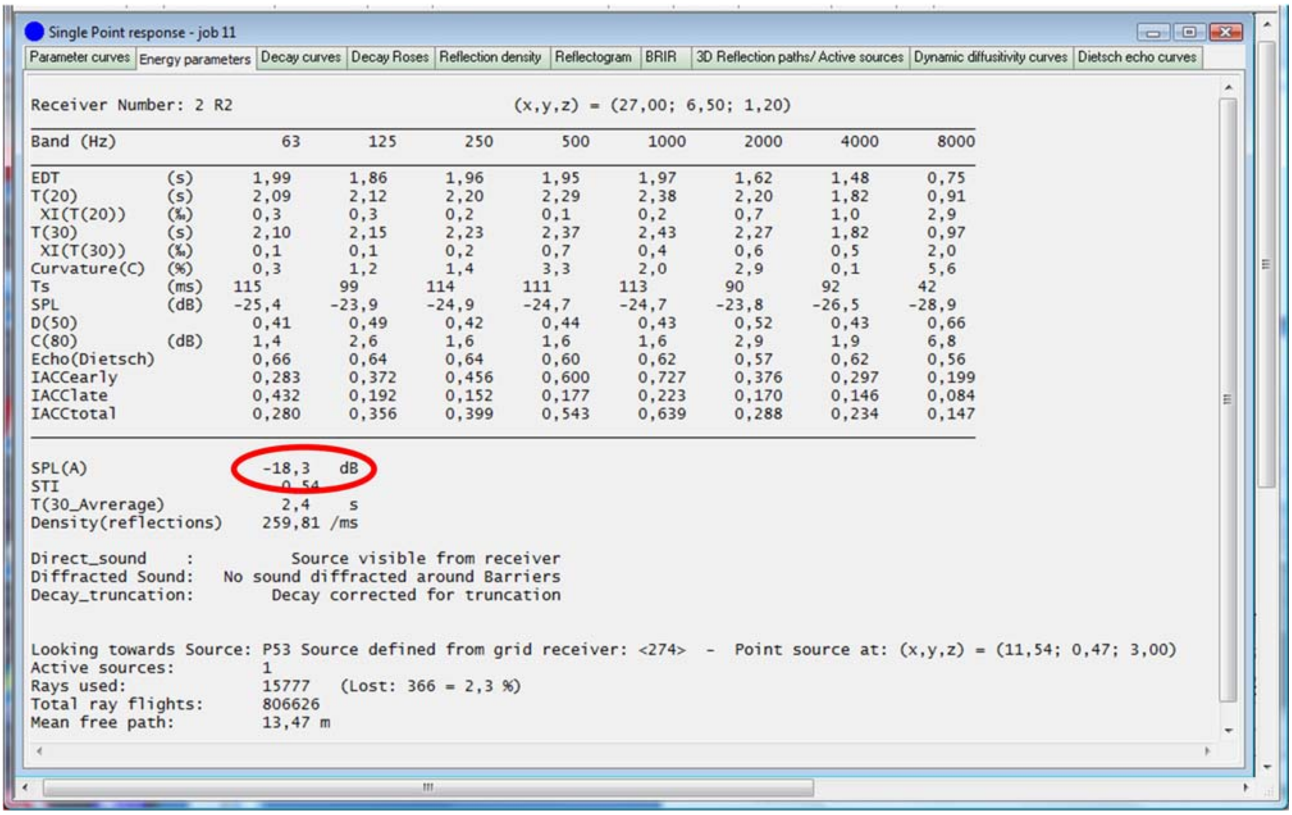
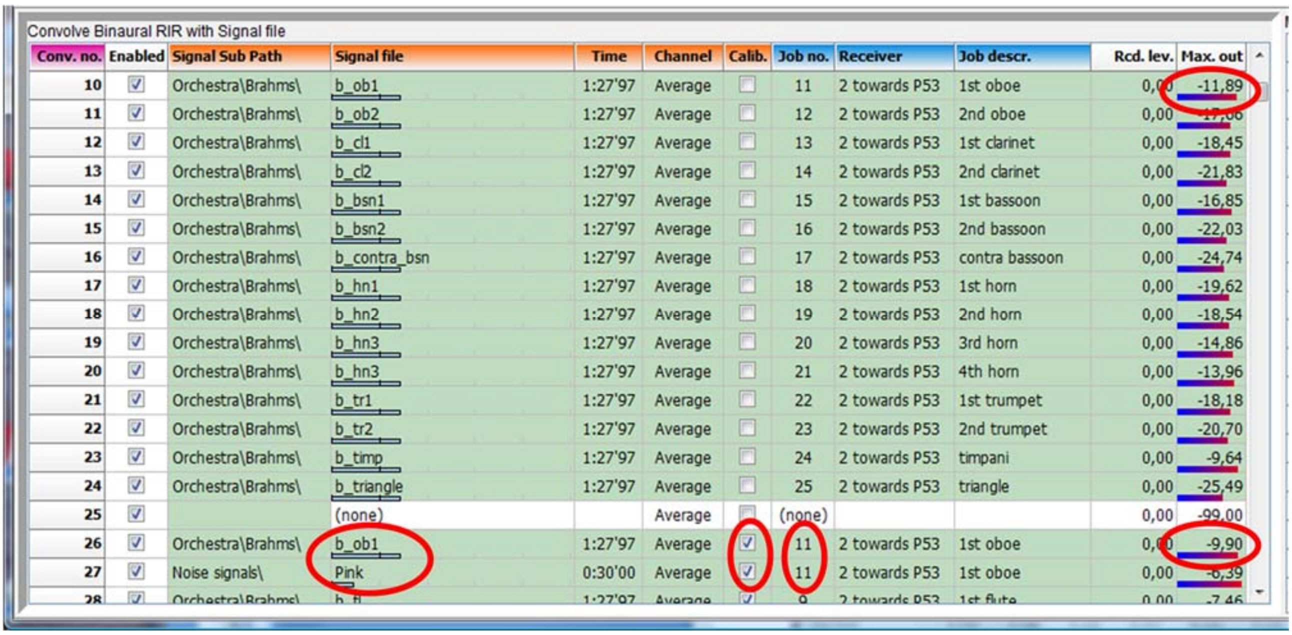


Figure 11. Source definition. The sound power level is $L_{W,A,1}$ = 7.0 dB(A).



Band (Hz)		63	125	250	500	1000	2000	4000	8000
EDT	(s)	1,99	1,86	1,96	1,95	1,97	1,62	1,48	0,75
T(20)	(s)	2,09	2,12	2,20	2,29	2,38	2,20	1,82	0,91
XI(T(20))	(‰)	0,3	0,3	0,2	0,1	0,2	0,7	1,0	2,9
T(30)	(s)	2,10	2,15	2,23	2,37	2,43	2,27	1,82	0,97
XI(T(30))	(‰)	0,1	0,1	0,2	0,7	0,4	0,6	0,5	2,0
Curvature(C)	(%)	0,3	1,2	1,4	3,3	2,0	2,9	0,1	5,6
Ts	(ms)	115	99	114	111	113	90	92	42
SPL	(dB)	-25,4	-23,9	-24,9	-24,7	-24,7	-23,8	-26,5	-28,9
D(50)		0,41	0,49	0,42	0,44	0,43	0,52	0,43	0,66
C(80)	(dB)	1,4	2,6	1,6	1,6	1,6	2,9	1,9	6,8
Echo(Dietsch)		0,66	0,64	0,64	0,60	0,62	0,57	0,62	0,56
IACCearly		0,283	0,372	0,456	0,600	0,727	0,376	0,297	0,199
IACClate		0,432	0,192	0,152	0,177	0,223	0,170	0,146	0,084
IACCtotal		0,280	0,356	0,399	0,543	0,639	0,288	0,234	0,147

Figure 12. Calculation result for Source P53 at receiver position R2. The sound pressure level is $L_{p,A,1}$ = -18.3 dB(A).



Conv. no.	Enabled	Signal Sub Path	Signal file	Time	Channel	Calib.	Job no.	Receiver	Job descr.	Rcd. lev.	Max. out
10	☑	Orchestra\Brahms\	b_ob1	1:27'97	Average	☐	11	2 towards P53	1st oboe	0,00	-11,89
11	☑	Orchestra\Brahms\	b_ob2	1:27'97	Average	☐	12	2 towards P53	2nd oboe	0,00	-17,06
12	☑	Orchestra\Brahms\	b_cl1	1:27'97	Average	☐	13	2 towards P53	1st clarinet	0,00	-18,45
13	☑	Orchestra\Brahms\	b_cl2	1:27'97	Average	☐	14	2 towards P53	2nd clarinet	0,00	-21,83
14	☑	Orchestra\Brahms\	b_bsn1	1:27'97	Average	☐	15	2 towards P53	1st bassoon	0,00	-16,85
15	☑	Orchestra\Brahms\	b_bsn2	1:27'97	Average	☐	16	2 towards P53	2nd bassoon	0,00	-22,03
16	☑	Orchestra\Brahms\	b_contra_bsn	1:27'97	Average	☐	17	2 towards P53	contra bassoon	0,00	-24,74
17	☑	Orchestra\Brahms\	b_hn1	1:27'97	Average	☐	18	2 towards P53	1st horn	0,00	-19,62
18	☑	Orchestra\Brahms\	b_hn2	1:27'97	Average	☐	19	2 towards P53	2nd horn	0,00	-18,54
19	☑	Orchestra\Brahms\	b_hn3	1:27'97	Average	☐	20	2 towards P53	3rd horn	0,00	-14,86
20	☑	Orchestra\Brahms\	b_hn3	1:27'97	Average	☐	21	2 towards P53	4th horn	0,00	-13,96
21	☑	Orchestra\Brahms\	b_tr1	1:27'97	Average	☐	22	2 towards P53	1st trumpet	0,00	-18,18
22	☑	Orchestra\Brahms\	b_tr2	1:27'97	Average	☐	23	2 towards P53	2nd trumpet	0,00	-20,70
23	☑	Orchestra\Brahms\	b_timp	1:27'97	Average	☐	24	2 towards P53	timpani	0,00	-9,64
24	☑	Orchestra\Brahms\	b_triangle	1:27'97	Average	☐	25	2 towards P53	triangle	0,00	-25,49
25	☑		(none)		Average	☐	(none)			0,00	-99,00
26	☑	Orchestra\Brahms\	b_ob1	1:27'97	Average	☑	11	2 towards P53	1st oboe	0,00	-9,90
27	☑	Noise signals\	Pink	0:30'00	Average	☑	11	2 towards P53	1st oboe	0,00	-6,39
28	☑	Orchestra\Brahms\	b_fl	1:27'97	Average	☑	9	2 towards P53	1st flute	0,00	-7,46

Figure 13. The convolution of the 1st oboe is repeated in no. 26 and 27, but with the *calib.* turned on.

In the orchestra simulation, all anechoic recordings of the various instruments have a certain sound level, which should not be changed. Otherwise, the balance between the instruments will be corrupted. This means that here we cannot use the *Calib.* setting as in the previous cases. Instead we make two copies of the line with convolution no. 10 (1st oboe), as shown in Figure 13. The convolution no. 26 is with the anechoic recording of the oboe, and no. 27 is with pink noise. Both of these convolutions are made with the *Calib.* turned on. We note the difference in the *Max. out* in convolutions no. 10 and 26: ΔL_{cal} = -9.9 + 11.9 dB = 2.0 dB. This means that the sound level is increased by 2 dB when the *Calib.* is turned on.

The level at the receiver position, as it should be when listening to the auralisation of the 1st oboe, can be calculated from the previous observations:

$L_{p,A,2} = L_{p,A,1} - L_{W,A,1} + L_{W,A,2}$ = - 18 - 7 + 94 dB = 69 dB.

Convolution no. 27 in Figure 13 is with the pink noise signal, and since the *Calib.* is turned on, this convolution has the same A-weighted SPL as no. 26 with the oboe. The A-weighted sound pressure level of these two convolutions is then calculated as:

$L_{p,A,2,cal} = L_{p,A,2} + \Delta L_{cal}$ = 69 + 2 = 71 dB.

If we want the calibration signal to be a round number like 80 dB, we have to add another 9 dB, which is shown in Mix. no. 6 in Figure 14. Note, that all other mixes must have zero *Mix. Lev.*

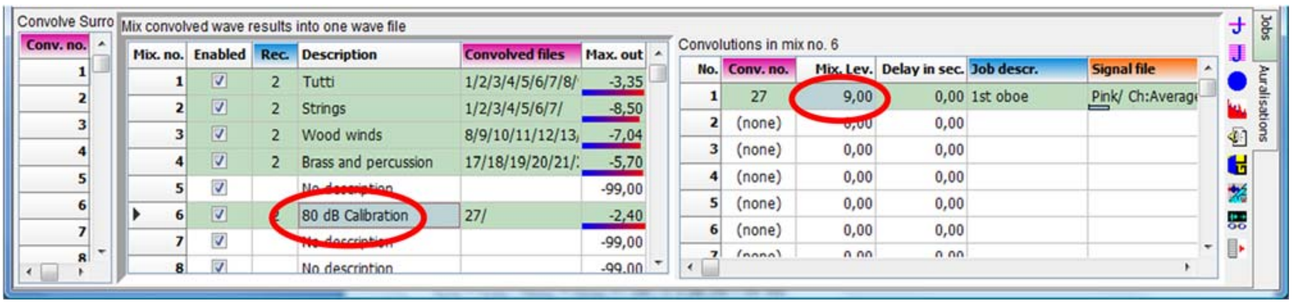


Figure 14. The auralisation mixer. Only for the calibration signal the *Mix. Lev.* is set different from 0 dB.

When auralizing these convolutions the pink noise signal (Mix. no. 6) is played first, and the level of the loudspeakers is adjusted to give 80 dB(A) at the listening position in the room. After that, the *tutti* orchestra (Mix. no. 1) and the other auralisation mixes can be presented with correct levels.

References

1. ANSI 3.5-1997. *Methods for Calculation of the Speech Intelligibility Index*. American National Standard. New York (Reaffirmed 2007).
2. ISO 9921: 2003. *Ergonomics – Assessment of speech communication.* International Organization for Standardization, Geneva, 2003.
3. 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, 2012.
4. ODEON Application Note: *Acoustics in Restaurants*. 2012.
5. ODEON Application Note: *Orchestra simulation and auralisation.* 2015.
6. J. Meyer: *Acoustics and the Performance of Music*. 5th Edition. Springer, 2009.