EuroRegio2016
June 13-15, 2016
Porto · Portugal

Calibration of auralisation presentations through loudspeakers

Jens Holger Rindel, Claus Lynge Christensen

Odeon A/S, Scion-DTU, DK-2800 Kgs. Lyngby, Denmark.
jhr@odeon.dk

Abstract
The correct level of presentation is important for all kinds of auralisation in order to be as realistic as possible. The method applied in ODEON 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 trick is to create modified versions of the input signals used for auralisation in such a way that the input signals have the same A-weighted 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.

Keywords: Auralisation, calibration, room acoustics.

PACS no. 43.55.Ka, 43.58.Vb

1 Introduction

The calibration method is described here as implemented in the ODEON room acoustics software. The method may be applied with any other room acoustics software under the condition that it is possible to create anechoic input signals that are modified to a certain, fixed A-weighted sound level.

2 Method

The calibration method described here is meant for loudspeaker presentation. 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. A sound level meter with A-weighting filter is needed for the calibration.
A pink noise signal is well suited for the purpose, because the sound pressure level is stable and not sensible to small variations of the position in a room. A tone signal, which is normally used for the

calibration of microphones, would be completely unsuited for this purpose, because of random variations within 25 dB or more in a room with approximately diffuse sound field.
In order to ensure a proper calibration, it is necessary to know the sound power level of the sound source to be simulated. In case of a talking person, the question is which vocal effort we want to simulate. In the case of a musical instrument, the question is at which dynamic level the music is played. Choosing the right sound power level in the simulation can be difficult and may be the major source of uncertainty in the auralisation 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)

Three cases will be described in the following with increasing degree of complexity.

3 Case 1: One person speaking

The following section describes the calibration of a single speech signal that simulates a person talking in a room. Speech levels can be taken either from ANSI 3.5 [1] or from ISO 9921 [2]. ANSI 3.5 has four levels of vocal effort from normal to shouted, see Table 1. The original data are given as sound pressure level (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 from SPL to sound power levels is further described in [3].

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

	Frequency, Hz								
Vocal effort	63	125	250	500	1000	2000	4000	8000	dBA
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**

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
$L_{W,A}$ (dB)	63	69	75	81	87

ISO 9921 [2] defines six 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, $L_{S,A,1m}$. Approximately 9 dB should be added in order to convert these data into A-weighted sound power levels.
A point source representing a talking person is selected for the simulation. The vocal effort is assumed to be *raised* and the spectrum and sound power level is 75.5 dB(A) in accordance with Table 1. The source is defined with 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
A simulation is made with a receiver position in the room model, and the calculated A-weighted sound pressure level is 54.8 dB ≈ 55 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*.

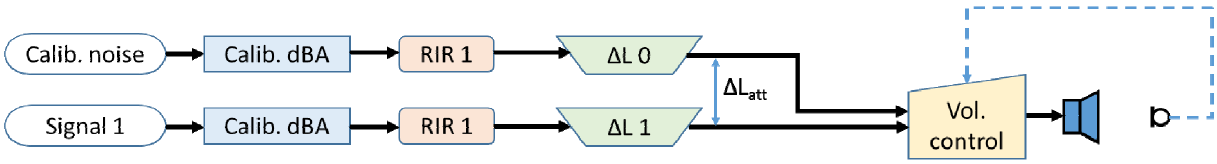

Figure 1. Diagram for the calibration process in case 1.

The calibration process is illustrated schematically in Figure 1. Two input signals are used; a calibration signal which is a pink noise signal with duration 30 s, and an anechoic recording of speech. The first step in the process is that both input signals are adjusted to have the same average level in dBA. Next step is the convolution with the room impulse response (RIR) calculated for the actual source-receiver positions. In this case, the same RIR is used for both signals, but it is possible to adjust the level of each signal before the power amplifier feeding the loudspeakers for the presentation.
The playback of the calibration signal should be sufficiently loud to make the calibration easy. Here we choose the A-weighted sound pressure level of the calibration signal $L_{p,A,2,cal}$ = 80 dB. This means that the calibration signal must be ΔL_{att} = 80 – 55 = 25 dB above the speech signal. 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 ΔL_0 = 10 dB, while the speech signal is adjusted down by ΔL_1 = -15 dB, see Figure 2.
When the *calib.* is checked (see Figure 2), ODEON will 'calibrate' the level of the input signal to have a fixed A-weighted level; so with two different recordings, here voice and noise, their loudness will be comparable. This is true also after the convolution if the same RIR is used for both signals.
Now the sound system should be turned on. The convolved pink noise (convolution no. 2) is then played, and the volume control is adjusted in order to get 80 dBA at the listening position in the room where the auralisation will be presented. When done, the calibration is finished and the sound system is ready for the presentation (convolution no. 1). The average A-weighted SPL should be close to 55 dBA 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, e.g. as convolution no. 4 in Figure 2. It is even possible to choose another anechoic recording, and the calibration is still valid. The calibration may also be valid for other sources and other rooms, as long as the level settings are the same. If for instance the point source is changed to a source with 7 dB higher sound power lever for *loud* vocal effort, the SPL of the auralisation is also increased by 7 dB. There is no need for a new calibration.

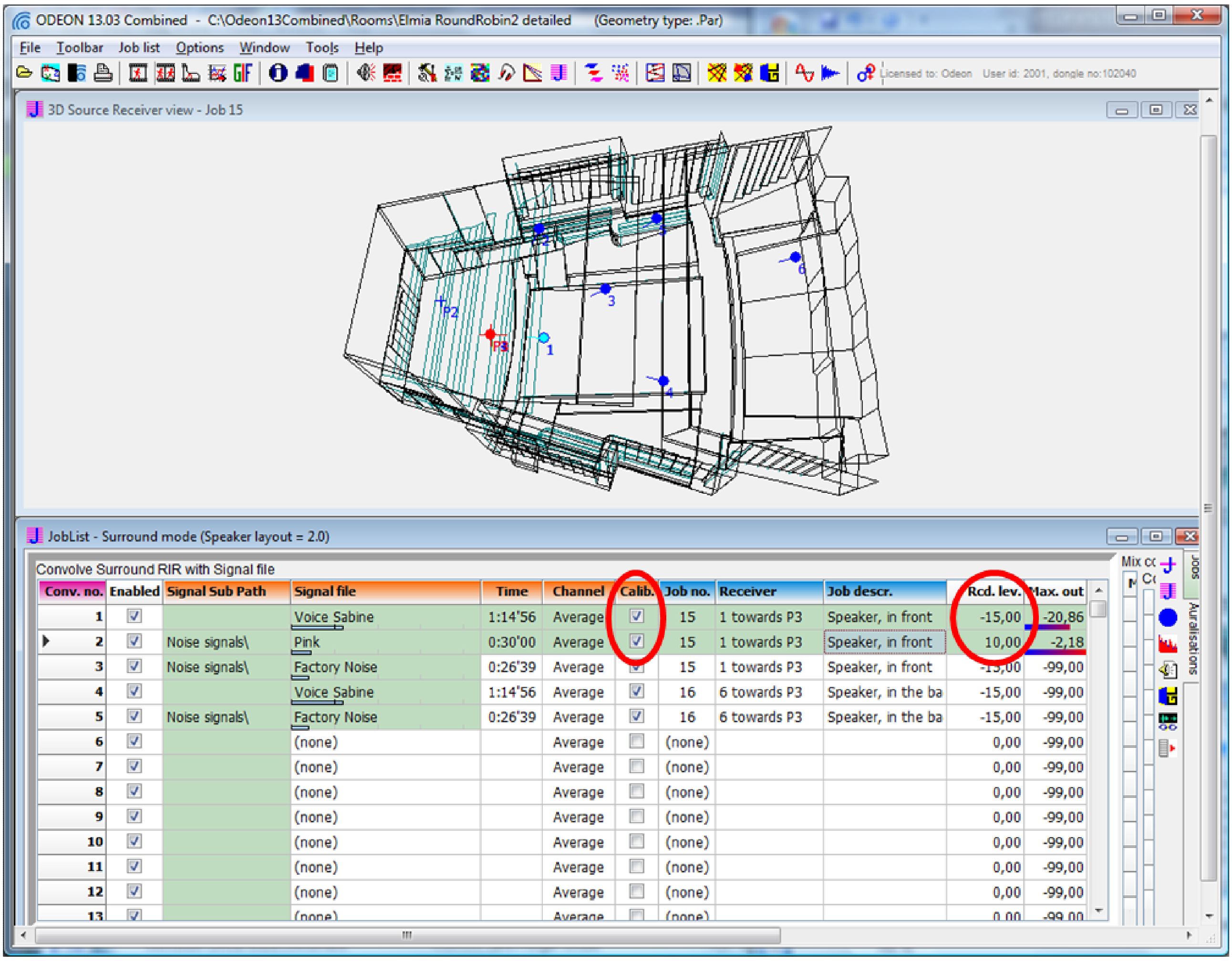

Figure 2. The auralisation page in the ODEON simulation. Convolution 1 and 2 are with speech and noise signal, respectively, and both using the impulse response from job 15. Note, that the Calib. is checked in order to adjust the input levels.

4 Case 2: Speech and background noise

Next case is the auralisation of verbal communication in 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 on modelling the acoustics of restaurants, see [3].
In this case the aim is to auralize different quality levels of verbal communication (see Table 3). The highest levels to be presented are speech at 74 dB combined with ambient noise at 83 dB.

Table 3. The sound pressure levels of speech and ambient noise and the associated quality of verbal communication [4].

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

The calibration process is illustrated schematically in Figure 3. Three input signals are used; a pink noise calibration signal, an anechoic recording of speech (signal 1), and a recording of a noisy party (signal 2). The receiver position is fixed, but different sources are used for the two signals, a point source for the conversation (RIR 1) and a surface source representing a noisy crowd in the restaurant (RIR 2). Thus, there is a level difference ΔL of the convoluted signals, which must be compensated for, i.e. $\Delta L = L_2 - L_1 = \Delta L_1 - \Delta L_2$.

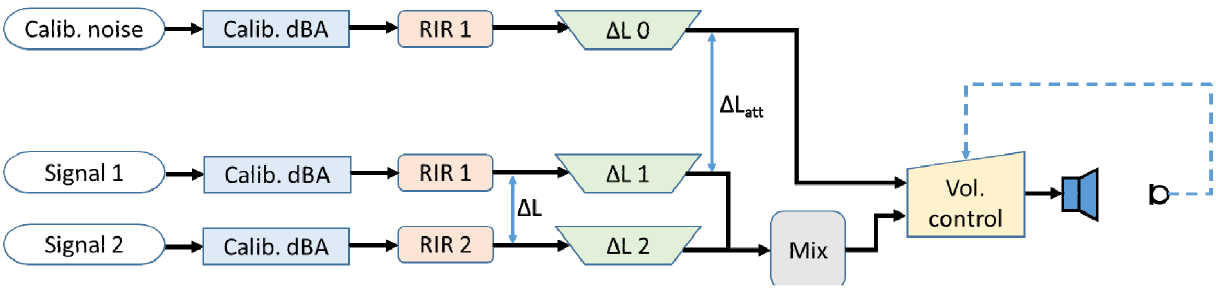

Figure 3. Diagram for the calibration process in case 2.

In Figure 4 is seen the auralisation page for the actual room simulation. Job 6 is with the surface source and job 7 is with the point source at 1 m distance. The *Calib.* is checked in all cases, which means that the relative difference in SPL is controlled by adjusting the level of each convolution. The A-weighted sound pressure level at the receiver point is L_2 = 81.0 dB in job 6 and L_1 = 68.1 dB in job 7. Thus, there is a difference of ΔL = 13 dB, which must be compensated for. In order to optimize the use of the dynamic range of the sound card, the talker convolution (job 7) is increased by ΔL_1 = 3 dB, and the ambient noise (job 7) is decreased by ΔL_2 = -10 dB.

In the right side of Figure 4 is seen the level setting of mix no. 1 having ambient noise signal adjusted by 0 dB, i.e. to create 83 dB; the speech signal adjusted by -9 dB, i.e. to create 74 dB. This is the case with SNR = -9 dB, which is the bottom row in Table 3. Similarly, other mixed signals can be prepared with the relevant level adjustments.

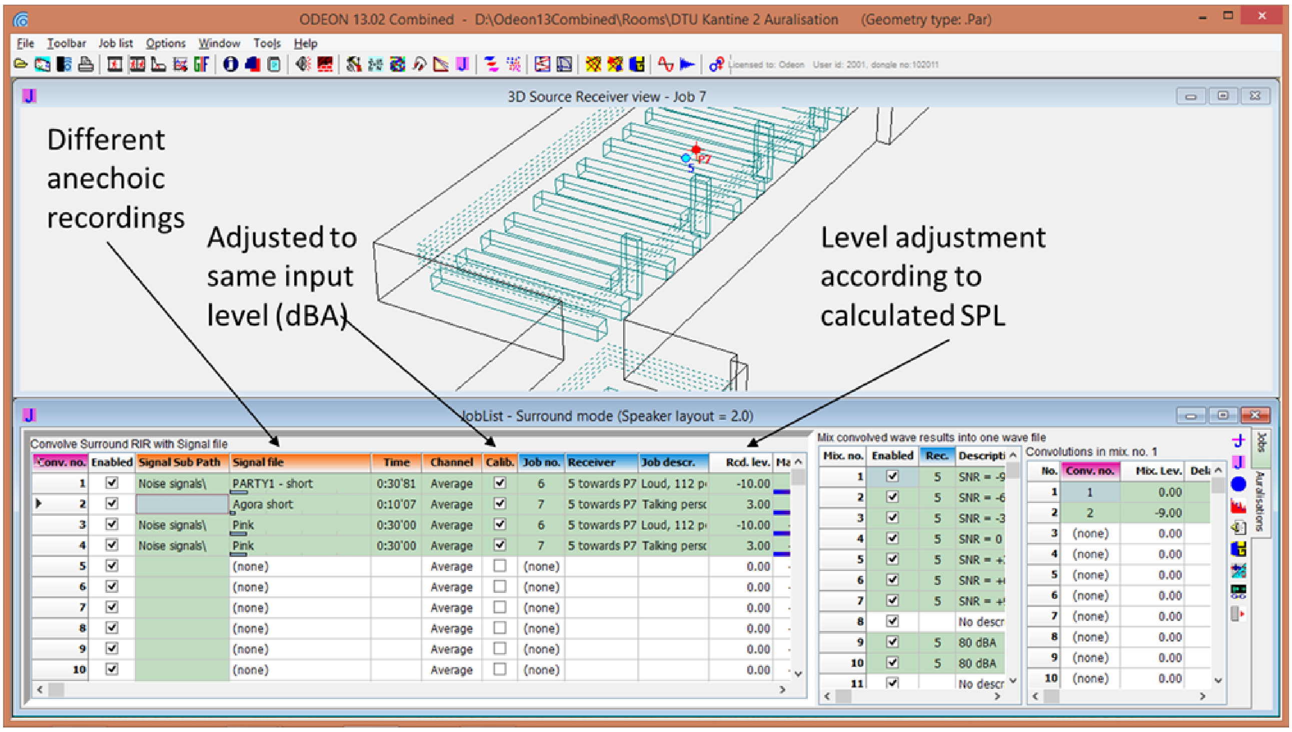

Figure 4. The auralisation page. Convolution 1 is with crowd noise, convolution 2 is with speech and convolutions 3 and 4 are with pink noise signal. Note, that the *Calib.* is activated.

For the calibration of these convolutions the pink noise signal (Conv. no. 4 in Figure 4) is played, and the level of the loudspeakers is adjusted to give 83 dBA at the listening position in the room. After that, the auralisation mixes can be presented with correct levels. With reference to Figure 3, this means that $\Delta L_0 = \Delta L_1 = 3$ dB. Alternatively, we could set $\Delta L_0 = 0$ dB and the calibration signal should be adjusted to 80 dBA.

5 Case 3: 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 presented in the following. As an example we look at the Brahms 4th symphony, 3rd movement, setup in a model of the Concertgebouw concert hall. In total 67 musical instruments are modelled, each as a point source with appropriate directivity. However, all point sources have the same overall sound power level, which is quite arbitrarily set to $L_{W,A,1} = 7$ dB.

One very important difference between this case and the previous ones is, that here the anechoic recordings of each instrument are carefully adjusted to have the correct relative sound power. This means, that here we cannot apply the automatic adjustment to equal A-weighted level for all signals. Otherwise, the balance between the instruments would be corrupted. 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 5.

For the calibration we choose one of the instruments, the 1st oboe, which is positioned in the middle of the orchestra, source position P53. The chosen receiver position for the calibration is receiver no. 2 in the stalls.

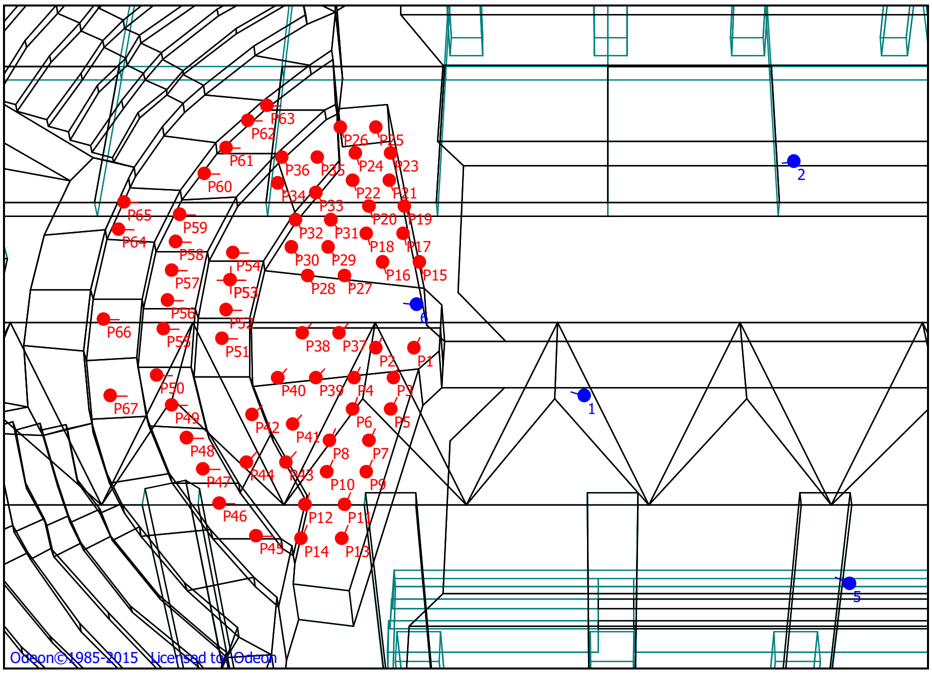

Figure 5. Orchestra setup in the concert hall model. The 1st oboe is source P53. Auralisation is made for receiver no. 2.

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. If no details are available about the actual music piece, forte level is assumed, i.e. $L_{W,A,2}$ = 93 dB.

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

However, a more detailed evaluation is possible in this case. If we look into the music score, we would notice that the dynamic notation is fortissimo (*ff*). 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 6, it appears that the oboe plays roughly only half of the time. This means minus 3 dB in the integrated sound level. Therefore, we end up with the sound power level for the oboe signal:

$$L_{W,A,2} = 93 + 4 - 3 \text{ dB} = 94 \text{ dB}.$$

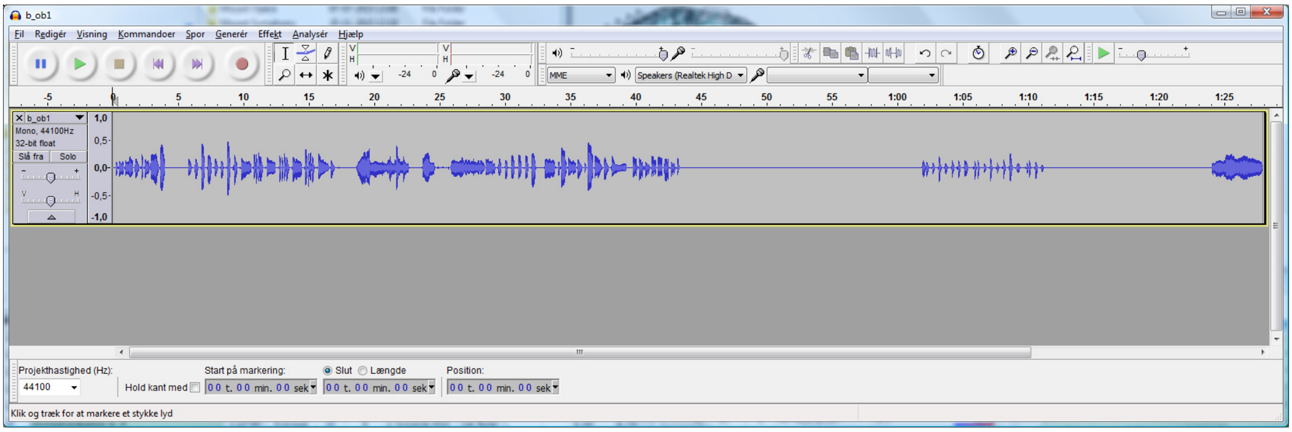

Figure 6. The sound file of the 1st oboe in the Brahms symphony anechoic recording.

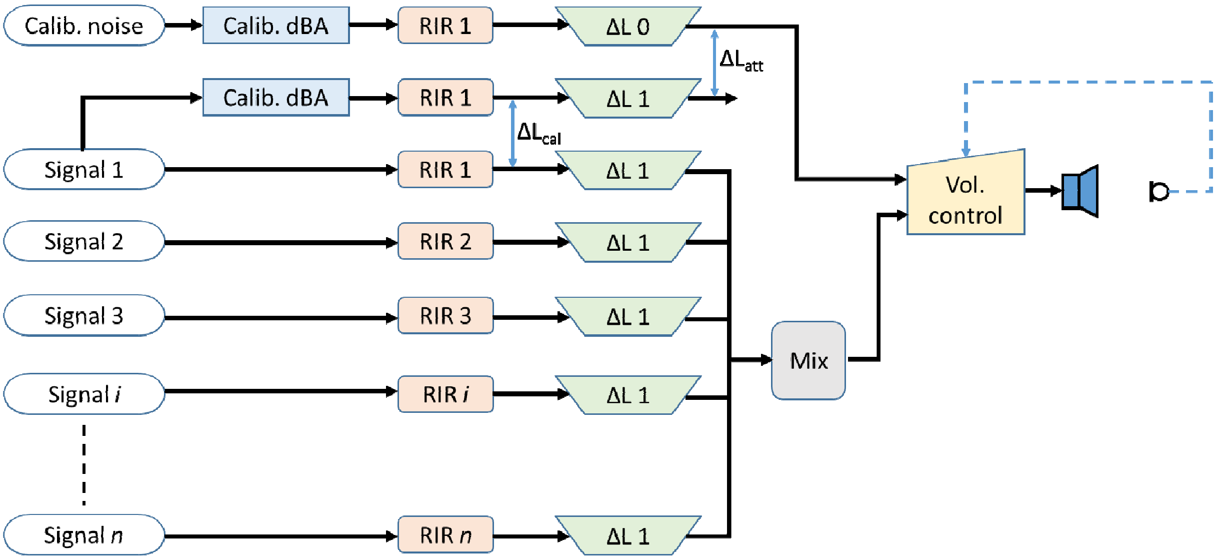

Figure 7. Diagram for the calibration process in case 3

The calibration process for the orchestra case is illustrated schematically in Figure 7. Signals 1 to *n* are the anechoic recordings of the musical instruments and the calibration signal is pink noise. The RIR is calculated for each instrument and used for the respective convolution.

As already mentioned, the sound power level of all the point sources is $L_{W,A,1}$ = 7.0 dB and for the 1st oboe the sound pressure level in point R2 is calculated as $L_{p,A,1}$ = -18.3 dB. Then the level $L_{p,A,2}$ at the receiver position, as it should be when listening to the auralisation of the 1st oboe, can be calculated:

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

In order to get a pink noise calibration signal, we use the RIR 1 for the 1st oboe as calculated for the source P53 to receiver R2 and include the A-weighting adjustment, see the upper line in the diagram, Figure 7. For comparison, we use the oboe anechoic recording (signal 1) both with and without the A-weighting adjustment, see the second and third line in the diagram. We note the difference in the *Max. out* in these convolutions: ΔL_{cal} = -9.9 + 11.9 dB = 2.0 dB (not shown here). This means that the sound level of the oboe signal is increased by 2 dB when the *Calib.* is turned on, and thus the level of the calibration signal is:

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

If we want the calibration signal to be a round number like 80 dB, we just have to add another $\Delta L_0 = 9$ dB. For all the other convolutions it is important that the same $\Delta L_1 = 0$ dB is applied. The exception could be that we want to change the balance between the instruments.

When listening to these auralisations, the pink noise signal is played first, and the level of the loudspeakers is adjusted to give 80 dBA at the actual listening position in the room used for the presentations. After that, the *tutti* orchestra or selected groups of instruments can be presented with correct levels. In the room model, listening position can be changed, sources can be moved, the room can be modified with other materials etc. New calculations can be made and the modified orchestra auralisations can be presented for comparison without calling for a new calibration.

6 Summary

When listening to auralisations, the sound level is very important for the realism. A method is suggested for calibration of loudspeaker presentations of auralisations using a pink noise signal as a calibration signal. In some cases it is possible to use a simple procedure, where the level of the anechoic signal is adjusted to have the same A-weighted level as the calibration signal. In other cases like the orchestra simulations, the level of the anechoic signals should not be changed; the calibration is still possible but becomes more complicated.

The most important source of uncertainty is determination of the correct sound power level of the source. In the case of speech, it is a matter of what vocal effort is correct. Standards from ISO and ANSI can be a good help. In the case of musical instruments, it is a matter of instrument type, dynamic level in the music, occurrence of pauses etc. To do this correctly may require some knowledge about musical instruments and the performance of music.

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] Rindel, J.H.; Christensen, C.L.; Gade, A.C.; Dynamic sound source for simulating the Lombard effect in room acoustic modeling software. *Proceedings of Internoise 2012*, New York, 2012.

[4] Rindel, J.H.; The acoustic of places for social gatherings. (Keynote). *Proceedings of EuroNoise 2015*, Maastricht, the Netherlands, June 2015, pp 2429-2436.

[5] ODEON Application Note: *Orchestra simulation and auralisation.* 2015.

[6] Meyer, J; *Acoustics and the Performance of Music*. 5th Edition. Springer, 2009.

Annex. Definition of symbols

$L_{S,A,1m}$	A-weighted sound pressure level of a talking person in the distance of 1 m in front of the mouth.
$L_{W,A,1}$	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.
ΔL_{i}	Adjustment of the level of signal no. *i* after the convolution.
ΔL_{att}	Offset of the calibration signal to achieve a certain calibration level.