# Dynamic sound source for simulating the Lombard effect in room acoustic modeling software

Jens Holger Rindel[a]
Claus Lynge Christensen[b]
Odeon A/S, Scion-DTU, Diplomvej 381, DK-2800 Kgs. Lynby, Denmark

Anders Christian Gade[c]
Gade & Mortensen Akustik, Hans Edvard Teglers Vej 5, 2. sal, DK-2920 Charlottenlund, Denmark

**The sound from speech when many people are gathered for a social event or in an eating establishment is a challenge for room acoustical modeling, because the sound power of the source varies with the level of the ambient noise, the so-called Lombard effect. A new method has been developed for modeling the ambient noise of many people speaking in a closed space. The sound source is a transparent surface placed just above the heads of people and radiating in all directions from a large number of randomly distributed points on the surface. The receivers are represented by a grid of points covering the same area as the source, but at the ears' height. The sound power and spectrum of the source are based on the ANSI 3.5 standard and adjusted taking the Lombard effect into account. Only two parameters are involved: the number of people speaking and the calculated transfer function from the surface source to the grid of receivers; the latter is strongly dependent on the reverberation time in the room. The method has been tested and verified by comparison to measured data in three cases with very different reverberation times and with 380, 480 and 530 people, respectively. Auralization can be made to demonstrate the difficulty of conversation in ambient noise.**

[a] email: jhr@odeon.dk
[b] email: clc@odeon.dk
[c] email: acg@gade-mortensen.dk

# 1  INTRODUCTION

It is well known that many people gathered in a restaurant or at a social function can be a very noisy event and conversation may only be possible with a raised voice level and at a close distance. For people with normal hearing it can be quite exhausting, and people using hearing aids may find that verbal communication is impossible.

Thus, the acoustics of restaurants and other eating facilities are very often unsatisfactory, but although it is a widespread problem, the literature dealing the possible solutions is very sparse. In order to deal with the problem two things are needed; guidelines on how to characterize satisfactory or insufficient acoustical conditions and a method for predicting the noise under typical conditions. Since the acoustical problems obviously depend strongly on the number or individuals present, typical conditions for the prediction could be set to full capacity or a certain fraction of full capacity. In an eating establishment full capacity means that the number of people equals the number of chairs. So, the furnishing and the arrangement of tables in the room are parameters that must be included in a prediction model.

The present paper suggests a prediction method that has been implemented in the room acoustics software ODEON. The method allows the prediction of the efficiency of various noise measures, e.g. treatment of some surfaces with sound absorbing material. It is also possible to auralize the result in order to demonstrate the difficulty of verbal communication in ambient noise, typical with a signal-to-noise level below 0 dB.

# 2  SPEECH IN NOISE

## 2.1  The Lombard effect

The vocal effort is characterized by the equivalent continuous A-weighted sound pressure level of the direct sound in front of a speaker in a distance of 1.0 m from the mouth. A description of the vocal effort in steps of 6 dB is given in ISO 9921 [1]. Normal vocal effort corresponds to a sound pressure level around 60 dB in the distance of 1 m.

It is a well known phenomenon that many people speaking in a room can create a high sound level, because the ambient noise from the other persons speaking means that everyone raises the voice, which again leads to a higher ambient noise level. This is called the Lombard effect after Étienne Lombard (1869 – 1920), who was the first to report that persons raised their voice when subjected to noise. The average relationship between speech level and ambient noise level is summarized in ISO 9921 [1].

The increase of the speech level as a function of the A-weighted ambient noise level is described by the rate $c$ (the Lombard slope). Lazarus [2] made a review of a large number of investigations, and he found that the Lombard slope could vary in the range $c = 0.5$ to $0.7$ dB/dB. The Lombard effect was found to start at an ambient noise level around 45 dB and a speech level of 55 dB. Assuming a linear relationship for noise levels above 45 dB, the speech level can be expressed in the equation:

$$L_{S,A,1m} = 55 + c \cdot (L_{N,A} - 45), \quad (\text{dB}) \tag{1}$$

where $L_{N,A}$ is the ambient noise level and $c$ is the Lombard slope. The valid range for this relationship is limited to speech levels above 55 dB or noise levels above 45 dB.

## 2.2 Simple prediction model for the ambient noise level

Applying simple assumptions concerning sound radiation and a diffuse sound field a calculation model for the ambient noise level was derived in [3]. By comparison with several independent cases of measured data covering a wide range of number of people present, it was found that only the Lombard slope $c = 0.5$ could make a reasonable agreement with the measured data. The same Lombard slope was found earlier by Gardner [4] based on results from several cases of dining rooms and social-hour type of assembly. The suggested prediction model [3] can be expressed in the simple equation:

$$L_{N,A} = 93 - 20\log\left(\frac{A}{N_S}\right)$$
$$= 93 - 20\log\left(\frac{A \cdot g}{N}\right), \text{ (dB)}$$

(2)

where $A$ is the equivalent absorption area (in m$^2$) of the room and $N_S$ is the number of simultaneously speaking persons. However, in general only the total number of people $N$ present in the room is known, and thus it is convenient to introduce the group size, defined as the average number of people per speaking person, $g = N / N_S$. The equivalent absorption area of the room should include the absorption from the persons themselves. The sound absorption per person depends on the clothing and position (standing, seated) and typical values are from 0.2 to 0.5 m$^2$. However, in a previous study the contribution of absorption from persons was found to be of minor importance [3].

A very interesting consequence of (2) is that the ambient noise level increases by 6 dB for each doubling of the number of individuals present. The same result was found by Gardner already in 1971 [4]. Another interesting consequence is an unusual strong influence of the amount of absorption in the room; doubling the absorption area leads to a 6 dB decrease of the ambient noise level.

## 2.3 Vocal effort and power spectra of speech

As a sound source the human voice covers a very wide dynamic range. In ANSI 3.5 [5] is found data for the level and spectrum for four ranges; normal speech, raised, loud, and shouted, see Table 1 and Fig. 1. The data are given as sound pressure levels in front of the mouth in a distance of 1 m. While this way of representing the source makes good sense for applications in telecommunication, it is not the preferred kind of source data for applications in room acoustics, where the sound power radiated by the source is more applicable. However, the conversion from SPL in a certain direction and distance to the sound power level is straight forward by applying the directional characteristics of a speaking person; the well documented data from ref. [6] has been used. Table 1 shows both the original octave band SPL 1 m in the frontal direction and the corresponding sound power spectra for each of the four levels of vocal effort.

## 3   METHOD

### 3.1  The surface source

When many people are speaking in a social gathering, the location of the sound sources is not well defined, because the same people are not speaking all the time. Statistically a certain

fraction of the persons are speaking, but which of the persons changes all the time. Thus a statistical approach has been chosen, and the source is modeled as a horizontal surface source in a certain height over the floor. In the simulation the sound is emitted from a large number of points randomly distributed over the surface; each point is the origin of ray tracing, and the rays are emitted in randomly determined directions. The radiation of the sound follows the Lambert cosine law with maximum radiation up and down.

A spherical radiation pattern has also been tried, but the Lambert radiation seems to give better results. The reason for that has not been studied yet, but a possible explanation might be, that all speaking persons are facing the tables, i.e. with maximum radiation inward and minimum radiation outward from the group of dining people.

## 3.2 Initial calculation of the room transfer function

In the initial calculation the sound power of the source is set to that of one single speaking person, $L_{W,A,1}$. The idea is to use the simulation model to create a room transfer function that is as close as possible to the actual sound field instead of relying on an assumption of a diffuse sound field. Thus the transfer function $F$ is here defined as:

$$F = L_{W,A,1} - L_{N,A,1} \approx 10\log\left(\frac{A}{4}\right), \text{ (dB)} \tag{3}$$

where $L_{N,A,1}$ is the A-weighted sound pressure level at the receiver. The approximation at the end is according to the classical diffuse field theory, where $A$ is the equivalent absorption area in m$^2$. However, in the following the diffuse field is not a necessary assumption, and we shall not apply the approximation.

Instead of a single receiver point, a grid of receivers covering the area with people is used, and the A-weighted sound pressure level is taken from the grid response as the 50% percentile. An example is shown in Fig. 2, where the receiver grid is 1.2 m above the floor and the surface source is 1.5 m above the floor.

## 3.3 Adjusting the sound power level of the source

With all people present and $N_S$ persons speaking simultaneously, the ambient noise level $L_{N,A}$ can be expressed by the same transfer function as for a single person (3), and thus:

$$L_{N,A} = L_{W,A,total} - F, \text{ (dB)} \tag{4}$$

where $L_{W,A,total}$ is the A-weighted sound power level emitted from all the speaking persons. From the simple prediction model (2) the ambient noise level can also be expressed using the transfer function $F$ instead of the assumption of a diffuse sound field:

$$L_{N,A} = 81 - 2 \cdot F + 20\log(N_S), \text{ (dB)} \tag{5}$$

In order to simulate the noise from $N_S$ speaking persons in the room, the sound power level of the sound source must be increase by the amount $\Delta L$:

$$\begin{aligned} \Delta L &= L_{W,A,total} - L_{W,A,1} \\ &= 81 + 20\log(N_S) + L_{N,A,1} - 2 \cdot L_{W,A,1}, \text{ (dB)} \end{aligned} \tag{6}$$

where the two last terms are known from the initial calculation, see above. What remains is only to insert the assumed number of speaking persons. A final calculation can be made with the sound source adjusted with a gain of $\Delta L$, but the result is already known to be:

$$L_{N,A} = L_{N,A,1} + \Delta L, \text{ (dB)} \tag{7}$$

Two observations can be made from equations (6) and (7), both characteristic for the Lombard effect. The first observation concerns the number of speaking persons, $N_S$; if this number is reduced to the half, $\Delta L$ and the ambient noise level $L_{N,A}$ will both be reduced by 6 dB, not by 3 dB as for other sound sources. Secondly, if the absorption in the room is changed e.g. to the double amount, the sound pressure level $L_{N,A,1}$ is reduced by 3 dB, which means that the ambient noise level $L_{N,A}$ will be reduced by 6 dB, not by 3 dB as for other sound sources. So, both the number of persons and the absorption area in the room have very strong influence on the ambient noise level.

## 3    THREE TEST CASES

Measurements were made at the Technical University of Denmark on occasion of the annual celebration on 6th May 2011. Several hundreds of people were dining in different halls, and three of them were selected for the measurements, see the plan in Fig. 3. The number of seats was 480, 530 and 360, respectively. Hall A had never before been used for a dining party, the ceiling height is 3.6 m, the surfaces are stone, concrete and glass and the mid-frequency reverberation time (with the tables) is 2.5 s. Hall B is a canteen with ceiling height 3.0 m and mid-frequency reverberation time 0.8 s. Hall C is a nearly square hall with glass walls , the ceiling height is 4.35 m and mid-frequency reverberation time 1.0 s. Photos from the three halls are shown in Fig. 4 together with views from the ODEON room models.

Reverberation times $T_{20}$ were measured with tables and chairs before the guests arrived. The integrated impulse response method was used with paper bags as impulsive noise sources and a sound level meter (B&K type 2260) for the measurements. The results in the three halls are shown in Fig. 5 as function of frequency. For the computer models the absorption coefficients of the materials have been adjusted in order to achieve approximately the same reverberation time as measured. The average of one source position and three microphone positions were used in all cases.

The A-weighted sound pressure levels were measured during the dining party using noise dose meters (B&K type 4445) suspended about 0.40 m under the ceiling. Three measurement positions were used in each hall as shown in Fig. 3. The A-weighted sound pressure levels were saved once a second and Fig. 6 shows the averaged results in each hall between 19:00 and 22:00. During the first half hour the noise level increases significantly (15 – 20 dB) but after that the level is relatively stable; two times, around 19:30 and again around 20:15, the noise level drops, probably when a main course has been served and people are more busy eating than talking. Analyzing the results in the two hours interval between 20:00 and 22:00 the A-weighted equivalent sound pressure levels were 87 dB in Hall A and 83 dB in Hall B and C.

The 1/1 octave band spectrum of the noise was measured during the evening using the integrated $L_{eq}$ values and measuring times around 2 to 5 minutes in each hall. A sound level meter (B&K type 2260) was used. The results are shown in Figs. 7 and 8.

## 4  CALCULATION RESULTS

The computer simulations were made according to the method outlined above in section 2, using a surface source 1.5 m above the floor and a receiver grid 1.2 m above the floor and a grid size of 2.0 m. The sound power level and spectrum of the noise source were as shown in Table 1; all four cases of vocal effort were used. The group size was set to $g = 3.5$ in accordance with previous experience [7]. The results of the grid responses had a statistical variation, and the 50% percentile was used, see Table 2. The range between the 10% and the 90% percentiles was 3 dB in Hall A and B, and 4 dB in Hall C. So, approximately the uncertainty of the results is ± 2 dB.

The results are shown in Table 2 and compared to the measured A-weighted sound pressure levels in the two hour interval 20:00 – 22:00. It is observed that the agreement is in general very close (less than 1 dB deviation), and the influence of the applied speech spectrum is very limited. So, as the A-weighted levels are concerned it is not so important which speech spectrum is applied.

However, looking at the results in octave bands, there is a significant difference between the spectra for different vocal effort. In Hall A and C the best agreement was found using the spectrum of 'loud' speech, see Fig. 7. In Hall B the noise level was slightly lower and the best agreement in the octave band values was found using the 'raised' speech spectrum, see Fig. 8. These results show that when a realistic speech spectrum is used, i.e. a spectrum that is in agreement with the expected vocal effort, the ODEON simulations give results very close to the measured octave band sound pressure levels.

In all three halls the noise level exceed the recommended maximum of 71 dB for sufficient quality of verbal communication, which has been defined by a signal-to-noise level SNR ≥ -3 dB [7]. Here a communication distance of 1.0 m has been assumed; of course communication in a shorter distance will give a better SNR. From (1) we derive a simple equation for predicting the SNR from a known ambient noise level:

$$SNR = L_{S,A,1m} - L_{N,A} = \tfrac{1}{2} \cdot (65 - L_{N,A}), \quad \text{(dB)} \tag{8}$$

As an example: Ambient noise level $L_{N,A} = 83$ dB, SNR = -9 dB and $L_{S,A,1m} = 74$ dB. This is the vocal effort called 'loud' speech in ANSI 3.5, see Table 3. This result is on the threshold between 'insufficient' and 'very bad' according the suggested rating in [7 and 8].

A suggestion for improvements may include treatment of the concrete ceiling in Hall A with a sound absorbing material. A new calculation was made in the ODEON model, using a material with absorption coefficients (125 – 4000 Hz): 0.45 – 0.55 – 0.60 – 0.90 – 0.86 – 0.75. The new calculation result was an A-weighted sound pressure level of 81 dB, i.e. a reduction of the ambient noise level by 6 dB. This example again demonstrates the unusual high efficiency of increasing sound absorption in the case of noise from speech compared to other noise sources.

## 5  AURALIZATION

In order to make auralizations of the simulated results, it is necessary to combine at least two signals, one representing a talking person in front of the listener in a distance of 1.0 m, and one or more other signals representing the ambient noise from many people speaking. The latter can be modeled by two point sources sufficiently far away from the listener, e.g. in a distance of 10 m. The relative level of the two auralized signals should be adjusted to achieve a mix with the

correct SNR as predicted from (8). Presenting the auralized result, the absolute level should be adjusted to be as close as possible to the predicted A-weighted sound pressure level.

## 6   DISCUSSION AND CONCLUSIONS

This paper suggests a method to model the dynamic sound source representing a large number of speaking people. It is a dynamic sound source because the sound power level depends on how much noise is created by the source in the room. There are no assumptions regarding the sound field (diffuse or other) as a room acoustics computer model is used to find the transfer function from a surface source to a receiver grid.

The spectrum of the speech signal depends on the vocal effort, but if only the A-weighted sound pressure level is considered, it is concluded that the dependence of the speech spectrum is negligible. However, if the octave band sound pressure levels are considered, it is important to apply a speech spectrum in agreement with the expected vocal effort. The comparison of simulated results with measured data from three acoustically very different halls has shown a surprisingly good agreement; the deviations of the simulated results from the A-weighted equivalent levels measured with two hours measuring time are less than 1 dB.

The most important and difficult parameter in the suggested method is the group size, i.e. the ratio between the number of individuals present and the number of speaking persons. In [3] it was found that the group size can vary between 2.5 and 9 in extreme cases, depending of the type of gathering. A small group size means a very lively party and vice versa. However, for a typical dining party it s found that a group size around 3.5 fits well with several cases of measured data.

## 7   REFERENCES

1.  ISO 9921. *Ergonomics – Assessment of speech communication*. (2003).

2.  H. Lazarus, "Prediction of Verbal Communication in Noise - A Review: Part 1", *Applied Acoustics* **19**, 439-464, (1986).

3.  J.H. Rindel, "Verbal communication and noise in eating establishments", *Applied Acoustics* **71**, 1156-1161, (2010).

4.  M.B. Gardner, "Factors Affecting Individual and Group Levels in Verbal Communication", *J. Audio. Eng. Soc.* **19**, 560-569, (1971).

5.  ANSI 3.5-1997. *American National Standard – Methods for Calculation of the Speech Intelligibility Index*, (1997).

6.  W.T. Chu, A.C.C. Warnock, *Detailed Directivity of Sound Fields Around Human Talkers*, IRC-RR 104, National Research Council, Canada (2002).

7.  J.H. Rindel: "Acoustical capacity as a means of noise control in eating establishments". *Proceedings of BNAM 2012*, Odense, Denmark, (2012).

8.  H. Lazarus, "Prediction of Verbal Communication in Noise - A Development of Generalized SIL Curves and the Quality of Communication (Part 2)", *Applied Acoustics* **20**, 245-261 (1987).

Table 1 – Speech spectra in octave bands for different vocal effort. SPL at 1 m in front of the mouth from ANSI 3.5 [5]. The sound power levels at 63 and 125 Hz are suggested based on measurements presented in this paper.

| 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 |

Table 2 – Results of calculations in three halls using different assumptions on vocal effort.

| Hall | Hall A | | | | | Hall B | | | | | Hall C | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Vocal effort | Normal | Raised | Loud | Shout | | Normal | Raised | Loud | Shout | | Normal | Raised | Loud | Shout |
| $L_{W,A,1}$ , dB | 68,4 | 75,5 | 82,6 | 91,0 | | 68,4 | 75,5 | 82,6 | 91,0 | | 68,4 | 75,5 | 82,6 | 91,0 |
| $L_{N,A,1}$ , (50%) dB | 50,5 | 57,6 | 64,6 | 73,0 | | 47,4 | 54,6 | 62,0 | 70,6 | | 48,9 | 56,1 | 63,3 | 71,9 |
| $N$ | 480 | 480 | 480 | 480 | | 530 | 530 | 530 | 530 | | 360 | 360 | 360 | 360 |
| $Ns$ ($g$ = 3.5) | 137 | 137 | 137 | 137 | | 151 | 151 | 151 | 151 | | 109 | 109 | 109 | 109 |
| $\Delta L$ , dB | *37,4* | *30,3* | *23,1* | *14,7* | | *35,2* | *28,2* | *21,4* | *13,2* | | *33,8* | *26,8* | *19,8* | *11,6* |
| $L_{N,A}$ , dB | 87,9 | 87,9 | 87,7 | 87,7 | | 82,6 | 82,8 | 83,4 | 83,8 | | 82,7 | 82,9 | 83,1 | 83,5 |
| $L_{A,eq,\ 2h}$ , dB | 87,3 | 87,3 | 87,3 | 87,3 | | 82,5 | 82,5 | 82,5 | 82,5 | | 82,9 | 82,9 | 82,9 | 82,9 |
| Deviation, dB | **0,6** | **0,6** | **0,4** | **0,4** | | **0,1** | **0,3** | **0,9** | **1,3** | | **-0,2** | **0,0** | **0,2** | **0,6** |

Table 3 – Quality of verbal communication and the relation to SNR (signal-to-noise ratio) [7, 8].

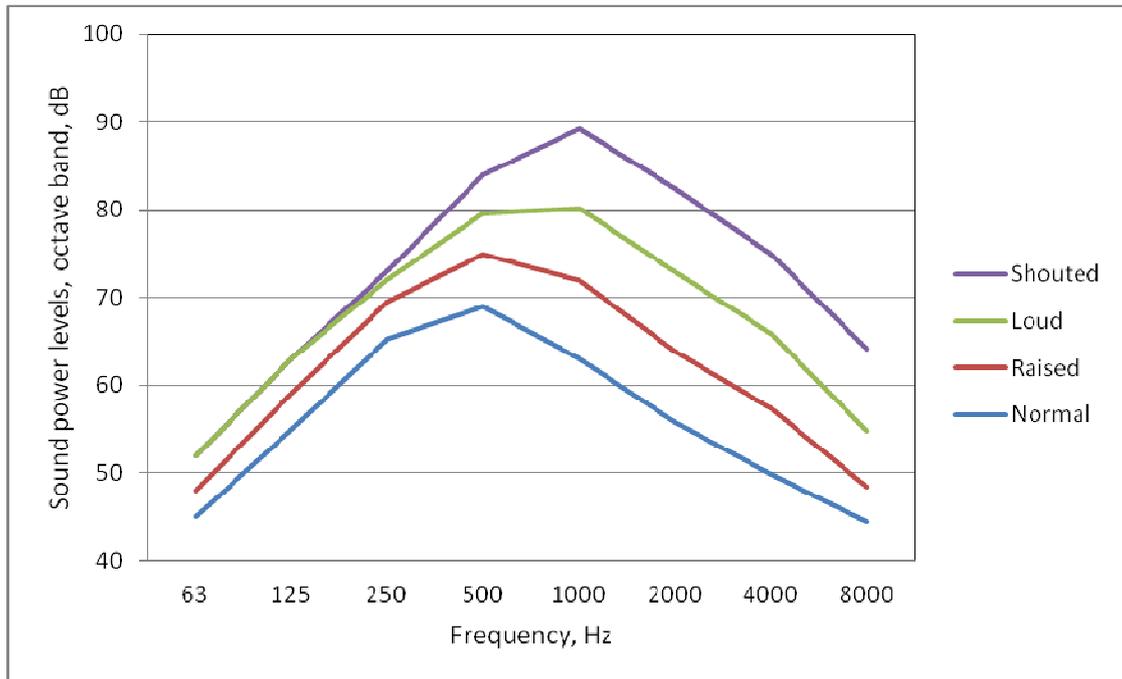| Quality of verbal communication | SNR dB | $L_{S,A,\ 1m}$ dB | $L_{NA}$ dB | $A/N$ m$^2$ |
|---|---|---|---|---|
| Very good | | | | |
| | 9 | 56 | 47 | (50 - 65) |
| Good | | | | |
| | 3 | 62 | 59 | (12 - 16) |
| Satisfactory | | | | |
| | 0 | 65 | 65 | (6 - 8) |
| Sufficient | | | | |
| | -3 | 68 | 71 | (3 - 4) |
| Insufficient | | | | |
| | -9 | 74 | 83 | (0.3 - 0.6) |
| Very bad | | | | |

*Fig. 1 – Speech spectra in octave bands for different vocal effort. Data at 250 Hz and above are from ANSI 3.5 [5], values below 250 Hz are suggested based on measurements presented in this paper.*
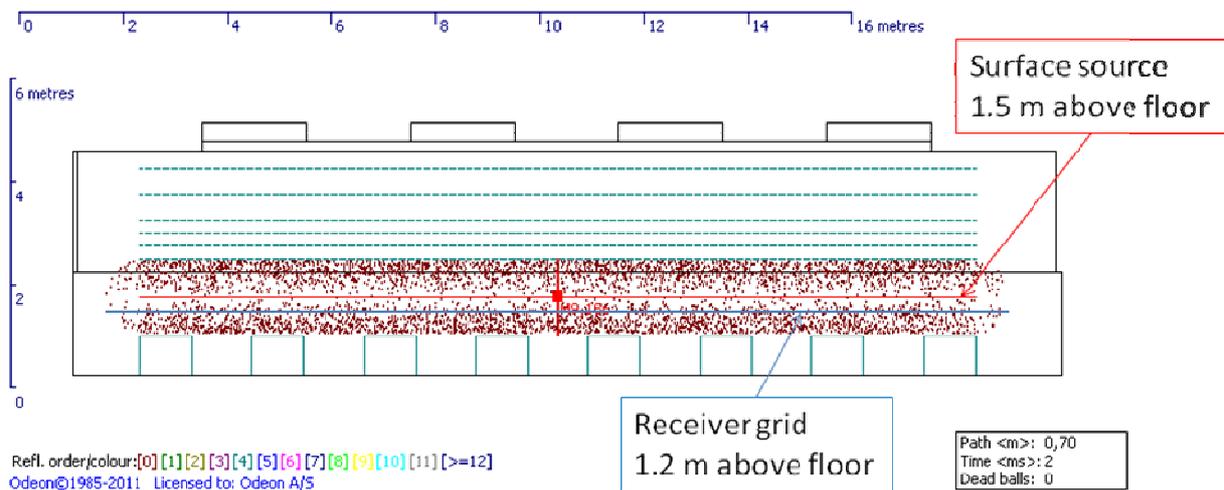


*Fig. 2 – Section of Hall C showing the position of surface source and receiver grid. The picture shows also emission of sound particles 2 ms after start.*

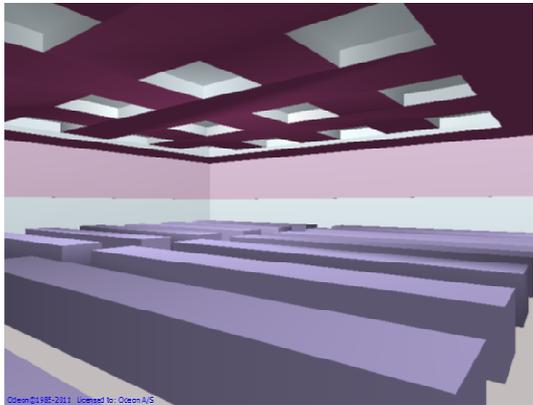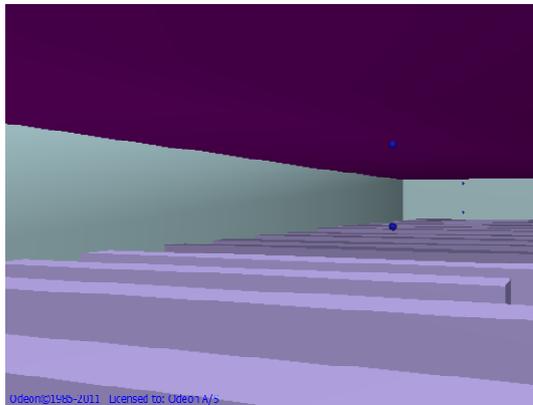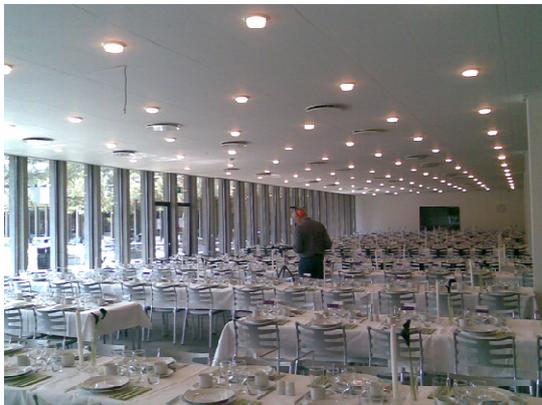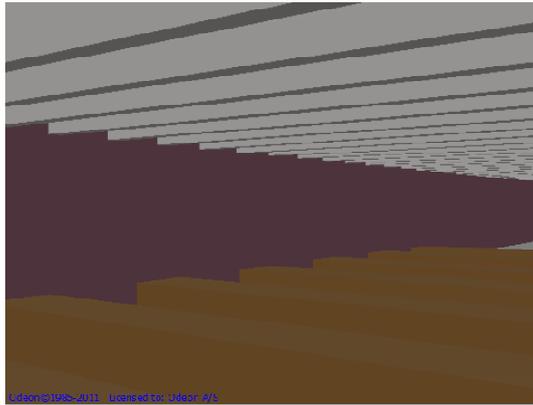*Fig. 3 – The three halls with table plans. The three microphone positions in each hall are marked in yellow.*

*Fig. 4 – Views from the three halls. Top: Hall A, Middle: Hall B, Bottom: Hall C. Photos from the real halls (left) and view from the computer models (right).*
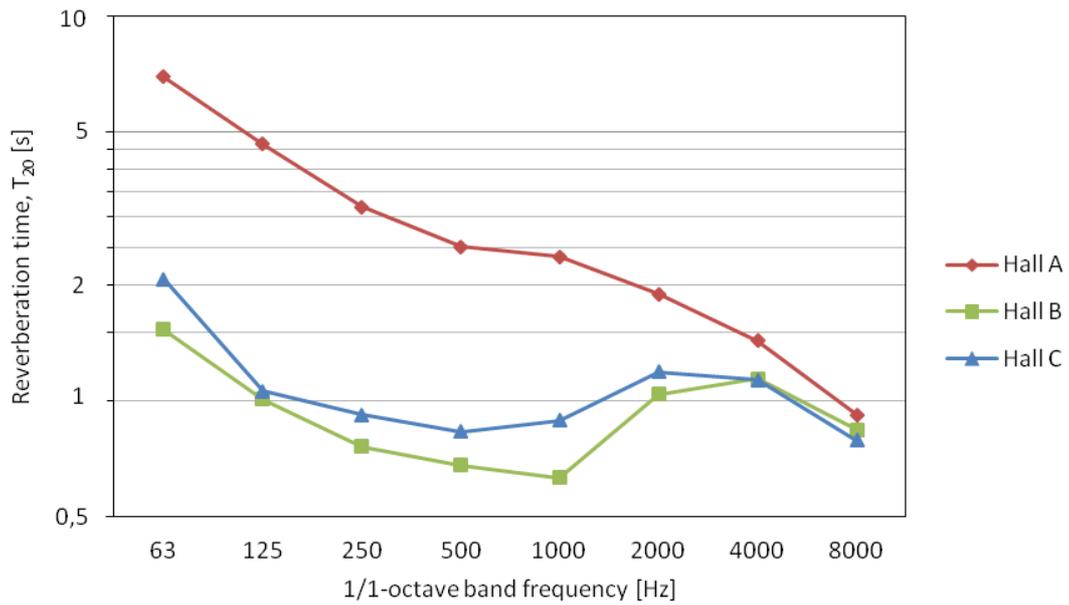
*Fig. 5 – Measured reverberation time $T_{20}$ in octave bands in each of the three halls.*
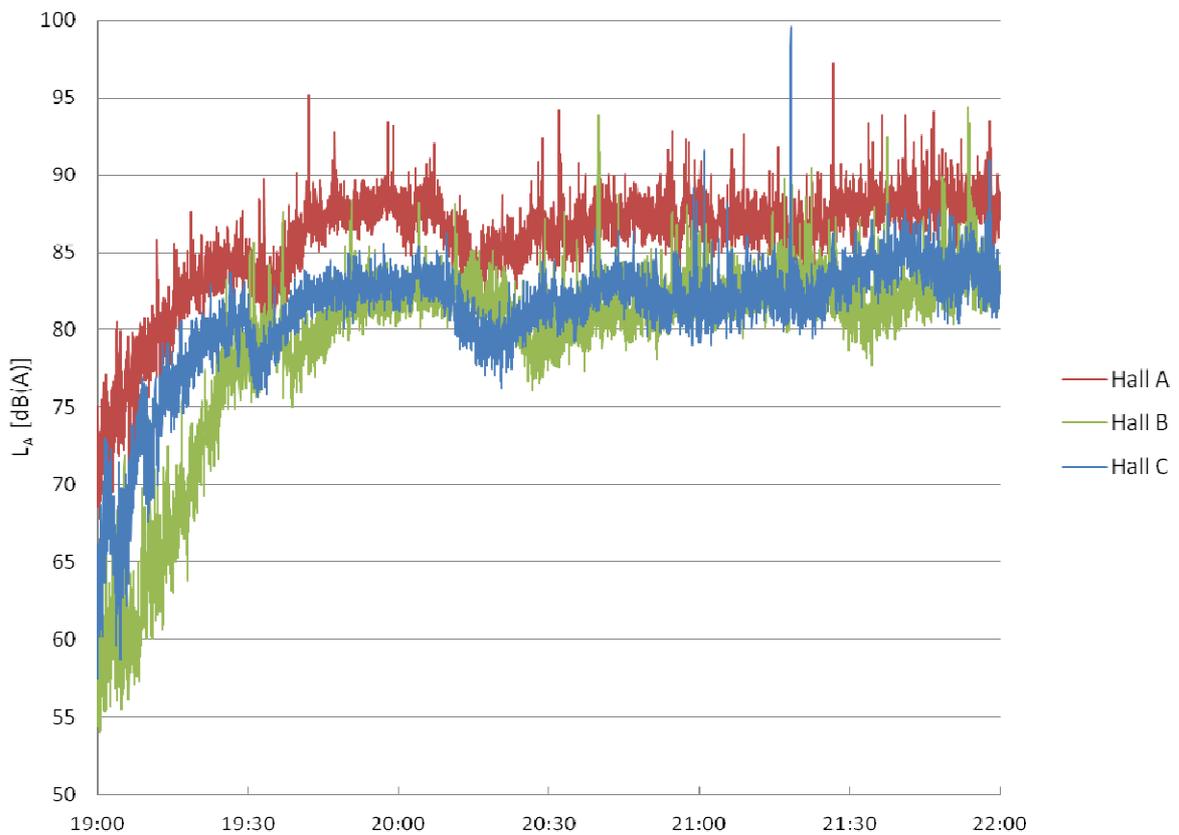


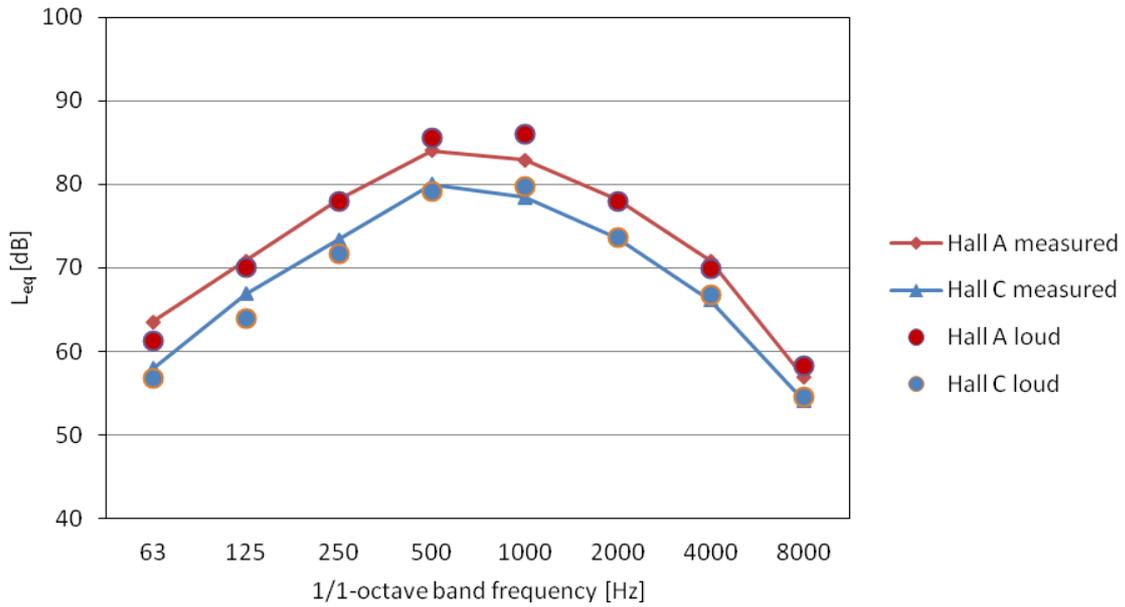*Fig. 6 – The measured A-weighted SPL during three hours in each of the three halls.*

*Fig. 7 – Measured and predicted noise spectra in hall A and C. Calculations were made with 'Loud' speech spectrum.*
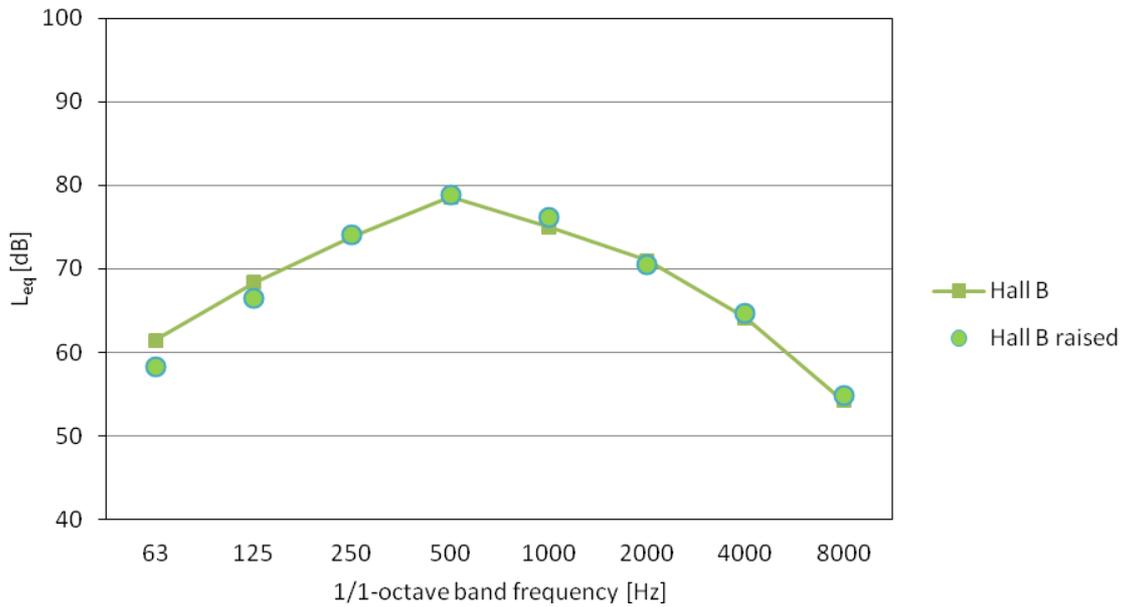


*Fig. 8 – Measured and predicted noise spectra in hall B. Calculations were made with 'Raised' speech spectrum.*