A Comparison of Room Simulation Software - The 2nd Round Robin on Room Acoustical Computer Simulation

Ingolf Bork
Physikalisch-Technische Bundesanstalt, Bundesallee 100, D-38116 Braunschweig, Germany

Summary
For the 2nd International Round Robin on Room Acoustical Computer Simulation, the acoustical properties of a modern Swedish concert hall were calculated by 16 participants and compared with the data measured by three teams. Within six octave bands the values of $T_{30}$, $EDT$, $D_{50}$, $C_{60}$, $G$, $TS$, $LF$, $LFC$ and $IACC$ had to be calculated using the geometrical data for the hall as given in maps, photos and descriptions of its surface properties. For a unique data input, a common geometry model and in the second phase the absorption and diffusivity data of the surfaces were given. The comparison of the results of the participants furnishes sources of calculation errors. First of all, the estimation of absorption data by the software user substantially affects the decay times calculated, therefore the acoustic experience of the user plays an important part. Even with uniform absorption and diffusivity data, the differences between the calculation results obtained by the participants can scarcely be attributed to individual software properties. The great error at low frequencies and the inability to handle edge diffraction effects today still limit the application to large rooms without obstructions. Since the accuracy of calculation is only one important point for the quality rating of room simulation software, the potential user of such software is provided with a list of criteria determining the qualification for a certain purpose.

PACS no. 43.55.Ka, 43.55.Mc, 43.58.Ta

1. Introduction
The ever increasing computing power of modern personal computers at decreasing costs today enables acousticians and architects to use sophisticated room simulation software for calculating the acoustic properties of all kinds of rooms. Not only do the properties of recent software releases include the calculation of sound distribution, various reverberation times, clarity and binaural quantities, but the auralization of the generated virtual sound fields has also become an integral part of these software packages. Room simulation software therefore increasingly replaces the use of scale models in the planning and improvement of acoustically relevant sites. The application of such software is not limited to computer modelling of rooms: due to its powerful auralization capabilities the use in sound recording industry as reverberation processor in conjunction with special convolution processors has allowed a quality to be achieved which is superior to that of usual studio reverberation units. According to a personal talk with the recording staff of a German classical recording company, today compact discs are available which have been mastered using a convolution processor in conjunction with a commercial room simulation software.

The ability to predict the acoustical properties of rooms has been demonstrated by various authors [1, 2, 3, 4, 5, 6] but their limitations and problems have also been documented [7, 8]. Data calculated and measured for some selected typical rooms are compared with more or less good agreement. The accuracy of calculation generally is hard to estimate and depends on numerous parameters not only inherent to the software. Therefore the most important question before purchasing a room simulation software is: which is the best one for a given application? But how should software quality be defined? To answer this question, several aspects of room simulation must be taken into account and will be discussed in the following.

In 1994, the First International Round Robin on Room Acoustical Simulation was launched in order to compare the calculation results of room acoustical parameters for a test room. The geometrical data of this room (the lecture hall of the Physikalisch-Technische Bundesanstalt in Braunschweig, Germany (PTB)) were forwarded to the participants together with photos, drawings and in a second phase a list of absorption coefficients [9]. The comparison of the results of 14 programs showed the variance of the calculations but on the other hand it served the participating software developers as feedback to improve the performance of their products. The positive responses to this Round Robin encouraged the PTB to launch another Round Robin with another test object which will be reported on here.

2. Room acoustical simulation software

2.1. Room acoustical parameters
To understand the influence of quality determining factors of room simulation software some fundamental principles have to be explained. The description of the acoustical properties of rooms today is based on standardized room acoustical parameters which are defined in ISO/DIS 3382 [10]. Since these parameters strongly depend on the location of the sound source $S$ and a receiver $R$, the results are valid only for a given source-receiver pair. In Table I the nine parameters are listed with a short description and a value for the subjective limen, the just noticeable difference (JND) at 1 kHz which is important for the assessment of calculated values.
Table 1. Room acoustical parameters according to ISO/DIS 3382 and their subjective limen at 1 kHz.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition (ISO/DIS 3382)</th>
<th>Subj. limen</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{50}/s$</td>
<td>Reverberation time, derived from $-5$ to $-35$ dB of the decay curve</td>
<td>5%</td>
</tr>
<tr>
<td>$EDT/s$</td>
<td>Early decay time, derived from 0 to $-10$ dB of the decay curve</td>
<td>5%</td>
</tr>
<tr>
<td>$D_{50}/%$</td>
<td>Clarity, early (0-80 ms) to late (80- ms - 200) energy ratio</td>
<td>5%</td>
</tr>
<tr>
<td>$C_{60}/$dB</td>
<td>Clarity, early (0-80 ms) to total energy ratio</td>
<td>1 dB</td>
</tr>
<tr>
<td>$T_0$ms</td>
<td>Centre time, time of 1st moment of energy impulse response</td>
<td>10 ms</td>
</tr>
<tr>
<td>$G$dB</td>
<td>Sound level related to omnidirectional free-field radiation at 10m distance</td>
<td>1 dB</td>
</tr>
<tr>
<td>$L_F/%$</td>
<td>Early lateral (5-80 ms) energy ratio, $\cos^2$ (lateral angle)</td>
<td>5%</td>
</tr>
<tr>
<td>$L_{FC}/%$</td>
<td>Early lateral (5-80 ms) energy ratio, cos(lateral angle)</td>
<td>5%</td>
</tr>
<tr>
<td>$IACC$</td>
<td>Interaural Cross Correlation Coefficient</td>
<td>0.2</td>
</tr>
</tbody>
</table>

These parameters describe the most important acoustical properties of rooms covering reverberation ($T_{50}$, $EDT$), clarity of speech ($D_{50}$) and music transmission ($C_{60}$), centre time $T_0$ and sound level of a source related to free field conditions $G$. Furthermore, some quantities describing the spatial impression are included: lateral energy fractions $L_F$ and $L_{FC}$, and the interaural cross correlation coefficient $IACC$.

The evaluation of all these parameters is based on the impulse response of the transmission between source and receiver, which can be measured by various methods among which the maximum length sequence technique (MLS) today is the most common one [11]. In the room simulation software the impulse response is calculated by image source and ray tracing algorithms which in some programs are based on statistical methods [7]. The evaluation of the parameters from the impulse responses in measurement and simulation is similar, therefore the comparison of calculated and measured impulse responses seems to be the first choice for the assessment of room simulation software. Otherwise, the numerical data of the room acoustical parameters are much better qualified for a software comparison since they show what a user gets out of the computer and what is needed for the interpretation of the simulation results. Furthermore, it enables the estimation of calculation errors and a graphical display including the results obtained by different softwares.

2.2. User interaction with the software

The results of a room acoustical computer simulation depend on many parameters that are not all accessible to the user. For a given test object, such as in our case a concert hall, the geometry data, i.e. the edge co-ordinates of all relevant surfaces and their absorption and diffusivity coefficients form the input for a black box “room simulation software” (Figure 1). Also, the locations of sources and receivers and their characteristics (directivity, power, frequency response) have to be defined by the user.

Depending on the software, some calculation parameters may also be set by the user, i.e. the number of rays (accuracy), reflection order, inclusion of diffuse reflections, length of the early, late and reverberant parts of the calculated impulse response, method to be used etc. The choice of all these parameters of course determines the computing time of the computer and the parameters therefore cannot simply be set to values furnishing the most accurate results. To get a result in a limited time, the user has to optimize these parameters. For the comparison of different softwares, this fact makes the result dependent on the skill of the user, and some experience is required to find the optimal values to be set. Further it is neither possible nor necessary to model each single edge in a concert hall so that correct results are obtained. It is up to the user to limit the number of surfaces without severe neglects, which otherwise may be limited by the capability of the software. A too high geometrical resolution could even reduce the accuracy of the calculation.

2.3. Quality aspects of room simulation software

The quality of room simulation software depends on many properties that may be of different importance to the individual user. Its usability is determined not only by the computing capacity but also by lots of additional features of the construction and debugging of the room model, as well as display and interpretation of the results. A very important point is the feedback between user and developer which should not be underestimated, since not all problems are described in the manuals and not all bugs are found by the programmer. Here a list of criteria and features for the comparison of room simulation software is proposed in order to help find the most suitable one:

- Calculation properties
  - Calculated parameters ($T_{30}$, $EDT$, $TS$, $C_{60}$, $D_{50}$, $G$, $LF$, $L_{FC}$, $IACC$, $RASTI$)
  - Accuracy of calculation, reliability of calculated results,
• Handling of special cases, e.g. open rooms, overlapping surfaces, curved surfaces, sound propagation over seats
• Inclusion of diffuse reflections and edge diffraction effects, dependence on incident angle
• Calculation with different receiver directivities, e.g. cardioid microphones, dummy head
• Control of calculation process by user
• Frequency range of calculations, results in octave bands, 1/3 octave bands
• Accuracy for low frequencies and small rooms
• Parameters of computing capacity controllable by the user
• User interface, usability
• Calculation speed

Geometrical model
• Geometry input tools
• Data conversion of geometry models from other CAD-platforms (DXF e.g.)
• 3-D display facilities, virtual walkthrough for checking the geometry and the surface properties (indicated by colours)
• Tools for model debugging

Libraries
• Library of applicable directivities and sound pressure levels in (1/3) octave bands of most common sound sources (loudspeakers, human voice, musical instruments)
• Library of headphone equalizations and head-related transfer functions (HRIR) of dummy heads
• Library of surface data (absorption and diffusivity) in octaves and 1/3 octaves

Presentation of results
• Quality and kind of displaying the results, parameter mapping
• Interpretation of results, e.g. display of ray paths

Auralization
• Auralization facilities, length of usable impulse response, sample rate (conversion), multiple sources with delays, binaural and multichannel features, surround sound, head tracking for binaural auralization, crosstalk cancelling of binaural signals for presentation via loudspeakers
• Interface to real-time convolution processors
• Inclusion of directivities of sources, with editing facilities

Support
• Documentation, handbook and on-line help
• Support by software developer (e-mail, Internet page)
• Contact with other users and exchange of experience

Software environment
• Operating system, hardware/software prerequisites (Windows NT, Win95/98, Linux, Apple/Macintosh)
• Stability of running software

All these aspects have to be kept in mind when a room simulation software is purchased. Since it is scarcely possible to get information about the accuracy of calculation of software, it was tried with the Round Robin, which is described in the following, to get an idea of the precision and the limits of calculating the acoustical properties of a concert hall.

2.4. Comparison of room simulation software

For a round robin on room simulation software, the aims must be defined. The objective comparison of software will be possible only if the input of each participating software is the same and all calculation parameters should be kept equal so that comparable results are obtained. Since the facilities for controlling the data processing in different software packages will vary, equal conditions are hard to achieve. Even if the co-ordinator of such a project could run all software packages on the same computer, it would not be possible to adjust all parameters in all softwares in the same manner. Furthermore, it takes a long time to get familiar with a software and to be able to obtain optimal results. Especially non-commercial research software requires special knowledge of the internal structure and can be used only by its developers. Therefore, also in this 2nd Round Robin, it was decided to give the input data to the participants and allow them to handle the data themselves.

The comparison of the calculation results will show only one, however important aspect of room simulation software: How good can the acoustical parameters of a selected room be predicted by such a software? What are the differences from the measured values? Since it was promised to the participants to keep the results anonymous, the Round Robin is not to be understood as a product test for a software. It should serve as a feedback for the programmers for the comparison of the calculation results with measured data of a concert hall which is scarcely accessible to everybody for individual measurements. Therefore the complete set of measurement data made available to the software developers enabled an improvement as a kind of fine tuning of their applied algorithms.

3. The 2nd Round Robin on room acoustical computer simulation

3.1. The test object

The ELMIA hall in Jönköping/Sweden was selected as test object. It is a modern multipurpose hall with variable acoustic elements on the side walls. These rotatable reflectors in their opened state reduce the reverberation of the hall by providing access to absorbing areas behind them. For measurement and simulation they were closed. The volume of the hall is about 11 000 m³ and it has 1100 seats. Figure 2 shows photos of the hall. For the calculation of impulse responses, two source positions on the stage, S1 and S2, were chosen and six receiver positions in the auditorium, R1 to R6. From the sketch in Figure 3 it can be seen that two receivers were located in the right hand balcony. Receiver R2 could not see source S2, and applying the rules fixed in ISO 3382 for source and receiver height of 1.5 m and 1.2 m respectively, the visual control at the time of measurement revealed that R5 could not see S2 either. It is shielded by the balustrade in front of the seats. This made the balcony seats more critical for the calculation as no direct sound could be taken into account and one program failed at calculating all the parameters for
the transmission from S2 to R2 because the direct sound path was not found. While in the first Round Robin only the 1kHz octave had to be calculated this time the most relevant six octave bands from 125 Hz to 4 kHz should be used.

3.2. Participants

Among the 16 participants from nine countries (Belgium, Denmark, Finland, France, Germany, Italy, Japan, Sweden, Switzerland) were not only the software developers but also software users. It is clear that these users do not have the same calculation facilities since they do not have access to the software source code (see Figure 1). Not all participants could take part in both phases of the Round Robin. Therefore the number of results in the graphs will be less than 16. The programs among which some are commercially available were: CATT, EPIDAURE, ODEON, RAMSETE, RAYNOISE, and also some research programs. Most of the programs apply combined image source/ray tracing algorithms, which today seems to be the optimal way to get the relevant data most quickly, including the early reflections. The effects of diffuse reflections are taken into account by nearly all programs. In the following the participants will be identified only by a number because they were guaranteed anonymity.

Since all participants could be contacted via e-mail and the result files forwarded as attachments the data exchange was very comfortable. For the handling of data, an Excel file was sent to all participants as a template into which the calculation data had to be entered and which then were sent back by the participants. It must be kept in mind that the calculation of 12 source-receiver-combinations for nine parameters in six octave bands furnishes 648 figures per participant to be entered into the computer. Therefore using copy and paste functions in a spreadsheet program were used to avoid typing errors in the data input.

3.3. Phases

3.3.1. Phase I

In accordance with Round Robin I [9], this Round Robin was divided into two main phases referred to as phase I and phase II. At the beginning of the Round Robin in October 1996, the material was distributed among the participants. It contained a number of photos, drawings and a description of the surface materials and the climatic conditions such as temperature (20°C), air pressure (1000 hPa) and humidity (50%). The positions of the reflectors and the movable elements above the stage as well as the locations of the sources and receivers were given. This phase is the usual case of an acoustic consulting situation, when a room has already been built and the properties of the walls and seats have to be estimated by the acoustician in order to calculate the actual acoustic properties of this room. In this phase, therefore, not only the quality of simulation software influences the results
but also the skill of the user as an acoustician. For estimating absorption coefficients and also the diffusivities of the given materials, it is not always possible to find suitable values in absorption tables, and only a few data of the diffusive reflection properties of the materials have been published until now. Also a lot of acoustic experience is necessary to know how detailed a geometrical model must be to provide a basis sufficient for the calculations.

3.3.2. The DXF model

This argument was the cause for another approach to the geometry modelling. Some participants who did not have the time to model the complete hall asked whether a computer model including all geometrical data would be available. The preferred format for such a model was that applied by the well-known CAD-software AutoCAD, which in some room simulation software can be used as input format for the import of geometrical data of rooms. For this kind of data, the file extension is DXF. It could indeed be managed to find a participant who offered his room model to all other participants on his web page. So a common room model was available to a group of participants, which enhanced the comparability of the software. Unfortunately, not every software enables data import via the DXF-interface, or some need AutoCAD software for conversion.

The identity numbers of all participants using this common room model will be extended by the letters DXF in the result graphs. For these participants, the number of unknown parameters for comparison of the simulation software will therefore be reduced. The differences in the calculations remain attributable to the following sources: the properties of the software itself, the surface data and the parameters accessible to the users (number of rays, number of reflections, order of diffusive reflections, time limit of early and late part of impulse response etc.).

3.3.3. Measurements

The measurements were carried out by the following three teams within two days (1997, July 1 and 2): Institute of Technical Acoustics (ITA) of the Technical University of Aachen; Department of Applied Acoustics, Chalmers University Göteborg; Section of Applied Acoustics, Physikalisch-Technische Bundesanstalt (PTB), Braunschweig. All teams used computer-based measuring systems using the MLS-measurement technique [11], strictly applying the rules fixed in ISO 3382. Mean values of these measurement data served as reference for the estimation of calculation errors. These data were sent to the participants at the beginning of phase II, when absorption and diffusivity data were distributed (February 1998).

3.3.4. Impulse response calculation

In order to separate two parts of the room simulation software another test was started: As shown in Figure 1, the calculation of the impulse responses is the central part of the evaluation process. The final results are obtained by applying integration, time windowing and filter algorithms to the impulse responses. It should therefore be proved that differences in the results should not be caused by these final calculations. In order to check the performance of the final data processing, one of the measured impulse responses (source S1, receiver R5, measured with the omnidirectional microphone) was distributed among the participants. They were asked to perform all calculations possible with a single impulse response, i.e. $T_{90}, E_{DT}, D_{50}, C_{60}$ and $T_{S}$. Unfortunately, only four participants were able to use their program for the evaluation of these parameters since at the corresponding processing stage this data input is not enabled by the program. One reason may be that the calculation of impulse responses is performed in octave bands and thus subsequent filtering is not applicable.

3.3.5. Phase II: with given absorption and diffusivity

In phase II of the Round Robin which started in February 1998, absorption and diffusivity data were distributed. Since measured data were not available for all the material used in the ELMIA hall, an estimation had to be carried out by the co-ordinator. It therefore could not be the aim of this phase to perfectly match the measured data by calculation, but rather to find out to what extent the different programs approach common values for the required parameters. In phase II the operator and software remained the only variables which could be responsible for the differences in the calculation results. Although the measurement data were at that time known to the participants, it is assumed that none of them did change his software algorithms in order to get closer to the measurements.

3.4. Results

The huge amount of data received from the participants made it impossible to show all results. Therefore only one excerpt of the most significant data will be given here in order to find out the most critical parameters and the variances by comparison of the different results of all participants. The participants received all data of all participants, anonymously assembled in a single spreadsheet file, for evaluation and comparison at the end of phase II of the Round Robin.

3.4.1. Evaluation of the impulse response S1R5

The evaluation of the measured unfiltered impulse response should show the accuracy of calculation in the last stage of the mathematical processing of the programs. Since this step is included in all programs, the final calculation results could never be more precise than the extraction of the acoustic parameter values from the impulse response. Therefore the standard deviation of these values helps to interpret the differences between the individual results of the participants. It should be kept in mind that the evaluation of all measured impulse responses is also affected by this uncertainty of calculation.

Figure 4 shows the five parameters which could be calculated with a single impulse response, the results of four participants (P1, P6, P8 and P9) and the corresponding measuring team (PTB) for the six octave bands. The corresponding standard deviations are listed in Table II. Comparison of
the curves in Figure 4 shows that the evaluation according to the rules fixed in ISO 3382 may leave arbitrariness in its application especially in the case of the early decay time $EDT$. Here the determination of the "initial 10 dB" may also be dependent on the filter applied and its steepness. The deviations of P6 may be due to a non-standard evaluation algorithm. The general time shift for the calculated centre time $TS$ for participant P1 seems to be a systematic error which may be due to an integration offset. If this error was compensated by a time shift of 40 ms, the standard deviations for $TS$ would be reduced to: 4.35 ms (125 Hz), 10.2 ms, 1.85 ms, 1.32 ms, 2.67 ms, 1.27 ms (4000 Hz) which seems to be a very good result. The rather great value for $\sigma(D)$ at 125 Hz corresponds to a very small calculated value of P9.
Table II: Standard deviations \( \sigma \) for five calculations of five parameters on impulse response S1R5 (cf. Figure 3).

<table>
<thead>
<tr>
<th>( f ) (Hz)</th>
<th>( \sigma(T_{30}) ) (s)</th>
<th>( \sigma(EDT) ) (s)</th>
<th>( \sigma(D) ) (%)</th>
<th>( \sigma(C) ) (dB)</th>
<th>( \sigma(TS) ) (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>0.030</td>
<td>0.093</td>
<td>4.65</td>
<td>0.28</td>
<td>13.97 (4.35)</td>
</tr>
<tr>
<td>250</td>
<td>0.044</td>
<td>0.144</td>
<td>2.52</td>
<td>0.36</td>
<td>20.90 (10.2)</td>
</tr>
<tr>
<td>500</td>
<td>0.024</td>
<td>0.011</td>
<td>1.41</td>
<td>0.15</td>
<td>18.18 (1.85)</td>
</tr>
<tr>
<td>1000</td>
<td>0.017</td>
<td>0.075</td>
<td>1.33</td>
<td>0.30</td>
<td>16.62 (1.32)</td>
</tr>
<tr>
<td>2000</td>
<td>0.019</td>
<td>0.022</td>
<td>1.61</td>
<td>0.18</td>
<td>17.20 (2.67)</td>
</tr>
<tr>
<td>4000</td>
<td>0.036</td>
<td>0.059</td>
<td>1.60</td>
<td>0.15</td>
<td>16.48 (1.27)</td>
</tr>
</tbody>
</table>

In Table III the standard deviations of the evaluated measurement on S1R5 performed by the three teams are given for comparison. It confirms that the error of P1 for \( TS \) is untypical of the accuracy of calculation.

This problem very clearly shows the need to also check the calculations performed in the programs, giving the software developer a better chance to improve the overall performance of the software.

3.4.2. Phase I: surface properties estimated

In phase I of the Round Robin, 14 participants sent their data based on their own geometry model or using the DXF model or in two cases (P5, P6) even both. Not all of the nine parameters in Table I could be calculated by all participants, so the \( IACC \) was in this phase calculated only by two participants. One participant (P12) carried out all calculations only in the 1 kHz octave, P10 did not calculate the 125 Hz octave.

To show the results, two ways were chosen:

- Display the data calculated for a parameter versus the source-receiver position in one octave band, to compare the change of a parameter with position.
- Display the data versus the centre frequency of the octave band at one source-receiver position, to compare the calculated frequency responses.

Since it is impossible to show all data here (12 positions and six octaves for nine parameters), only a few significant results can be discussed here. E.g. in Figure 5 the results for S1R5 are shown via the six octave bands. Note that the curves for the participants using the DXF model are dotted. The single value of participant P12 at 1 kHz is invisible here, since the calculated value (2.13 s) matches the measured curve. Since the variation of \( T_{30} \) at different positions is quite small, the dependence on frequency seems to be more significant here for displaying the accuracy of the calculation results.

In the mean, the data calculated by most participants lie around the measured data although the general tendency that the values decrease with frequency is not unique. In contrast to the measurement, some participants obtained too small values in the lower octaves. The two extreme curves (in both cases software users) show the width of the variance of calculations with room simulation software: at 500 Hz the \( T_{30} \)-values range from 0.9 s to 3.4 s. Astonishingly P15 furnished constant values for all frequencies at most positions.

The causes for these discrepancies may be due to the estimation of absorption parameters or rounding errors, whereas the geometry of the DXF model cannot be responsible. This shows that the skill of the user as an acoustician and also as a software user plays an important role for getting correct results using room simulation software. On the other hand, it was found that although the reverberation data were not correct, other parameters of these participants are closer to the measured values than those of other participants.

The extremely low \( T_{30} \) values for P15 correspond to the absorption coefficients chosen for the seats which seem much too high: 0.8 (125 Hz), 0.80, 0.85, 0.90, 0.92, 0.90 (4000 Hz). The values of P1 who obtained the best approximation to the measured \( T_{30} \) values were: 0.10 (125 Hz) 0.36 0.44 0.48 0.48 (4000 Hz). These differences clearly show the correlation between the absorption of the seats in the auditorium and the calculated decay times and therefore the dependence of the calculated results on the absorption estimated by the user. To get an idea of the variety of estimated absorption data for the seats in the main area of the auditorium, the extreme values used by the participants (as far as known) are given in Table IV.

Another way to display the individual deviation from the measured data is shown in Figure 6. Here for all participants the mean error for all positions in one octave band (1 kHz)
Table IV. Range of absorption data for the seats as estimated by the participants in phase I.

<table>
<thead>
<tr>
<th>[Hz]</th>
<th>125</th>
<th>250</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
</tr>
</thead>
<tbody>
<tr>
<td>max.</td>
<td>0.80</td>
<td>0.80</td>
<td>0.85</td>
<td>0.90</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td>min.</td>
<td>0.10</td>
<td>0.36</td>
<td>0.44</td>
<td>0.48</td>
<td>0.48</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Figure 6. Mean errors for the 1000 Hz octave over 12 positions and mean errors at S1R5 over six octaves for the participants in phase I.

and for all octaves in one position (S1R5) respectively is given to show the differences between the participants. The mean error is calculated by

$$\bar{e} = \frac{1}{N} \sum_{n=1}^{N} |x_c(n) - x_m(n)|,$$  \hspace{1cm} (1)

with $N =$ number of positions (12) or octaves (6), $x_m(n)$ - measured parameter, $x_c(n)$ - calculated value.

The comparison of the two mean values for the individual participants shows that some programs perform well along the frequency axis (low mean(octaves)-values) and others show better results along the position scale, in such cases the varying decay times of the individual source and receiver locations are well approximated. Although the reverberation time should not change very much with changing positions, the calculations have to consider different geometrical conditions and need not necessarily furnish constant values.

Thus Figure 6 shows a comparison of the performance quality of different programs indicating clearly the more accurate results by low mean error values. Here P1 calculated the best values from both aspects. For those participants who also used the DXF-model (P5 and P6), it can be seen that their own geometry model furnishes better results, which in the case of P5 contains only 94 surfaces instead of the 566 of the DXF Model. A comparison of the results of all DXF-participants shows that the geometry model alone is not responsible for errors in the calculated reverberation time $T_{30}$, i.e. no general tendency to greater errors can be seen. The great deviation of P13 and P15 are assumed not to be evoked by the model, because P7 got quite good results with the same geometry. Even a high number of surfaces alone will not lead to a better result: P9 in this phase used a model with 3530 surfaces while the other participants used from 94 to 566 surfaces.

Figure 7. Calculation results of $D_{30}$ at S1R1.

A more location-dependent parameter is the "definition" (or "Deutlichkeit") $D_{30}$: it represents the ratio of the energy in the first 50 ms to the total energy in the impulse response. Figures 7 and 8 show the calculation results at S1R1 and in the 125 Hz octave respectively. R1 is located at the centre seat in the 3rd row (compare Figure 3). Although this position very close to the source S1 seems at first sight to be uncritical, the results indicate a fundamental problem: most results are higher than the measured ones (plotted as thick solid line), and especially in the lowest octave (125 Hz) the measured data drops down to a value below 20. Since all three measuring teams found this low frequency value (a standard deviation of 2.51%), a measurement error can be excluded. It therefore is unclear why this position in the first 50 ms the energy is so low. A comparison of the $C_{30}$ values reveals the same effect in the first 80 ms. Figure 8 shows the results for this critical octave for all other source-receiver combinations. In most cases, the measured values are significantly greater than the calculated ones, except at position S2R1. This reflects the well-known problems of calculations restricted to geometrical acoustics in the low frequency range which may be due to grazing incidence; seat dip attenuation which presumably is neglected by most programs [12].
One possible explanation for the error at S1R1 may be the suppression of the phase in room simulation programs. The source loudspeaker S1 on the stage radiated at a height of 1.5 m and at a distance of 2 m from the edge of the stage. Geometrical considerations revealed that the receiver cannot "see" the image source below the stage surface, therefore a destructive interference effect can be excluded. Another explanation may be diffraction effects at the stage edge, where the sound from the source generates secondary sources obeying Huygen's principle, which have opposite phases due to the expansion of the wavefront and therefore tend to cancel the direct sound. This is enhanced by the fact that both source and receiver lie on the symmetry axis of the stage and the wavefront from the source reaches the slightly curved stage edge (see Figure 3) with equal phase. The wavelength in the 125 Hz octave ranges from 1.93 m to 3.86 m which may correspond to the distance between source and stage edge and also to the height of the source. This effect should be investigated in more detail by powerful sound field calculation programs (e.g. FEM/BEM) taking diffraction effects into account.

The results of all parameters of the 1st phase of Round Robin II are represented in the three graphs of Figure 9. In Figure 9a the two decay times $T_{30}$ and $EDT$ are displayed as mean errors averaged over six octaves and 12 positions (=72 values for each participant!), giving an overall estimation for the accuracy of calculation. Here and in the following graphs, the mean error is related to the subjective limen for the corresponding parameter. This has two advantages: 1. the errors of all parameters fit the same scaling, 2. it may be easily estimated whether or not the error is in an audible range. The reference values for the nine parameters are given in Table I. Although these values are valid for the 1 kHz octave only, they may here serve as a rough reference and for normalization.

When comparing the results of the participants it must be kept in mind that the calculation of mean values assumes that all participants have calculated all values. Those participants who did not calculate the most critical 125 Hz octave (P10 and P12) should therefore be excluded when the results are rated. For these participants the values shown in Figure 9 are averages over five (P10) or one octave (P12) respectively.

It should be remembered that all calculated results also depend on the participant's estimation of absorption and diffusivity factors and do not reflect the quality of the software alone but also the feeling of the operator for the surface properties.

The differences in the two decay time parameters $T_{30}$ and $EDT$ (Figure 9a) show, that the calculation is very critical in the early sound. While in most cases the $T_{30}$ values are more accurate than the corresponding $EDT$ values, only a few participants seem to use a good early sound approximation (e.g. P8 and P11) which in the mean is equal to or better than the $T_{30}$ calculation.

The comparison of the "clarity" parameters $D_{50}$ and $C_{80}$ (Figure 9b) shows almost equal relative errors, which is due to the similar calculation procedure. Note that in contrast to the first Round Robin [9] the subjective limen for $C_{80}$ has changed from 0.5 dB to 1 dB according to a most recent publication of measurements on the just noticeable difference of $C_{80}$ [13]. The similarity of the relative calculation errors for $C_{80}$ and $D_{50}$ confirms this change is reasonable.

A comparison of the TS and G values shows that there is no correlation between the error of G and that of TS. This means that although the surface properties and geometry were used in the individual programs, the accuracies for G and TS are different and depend only on the program. A program with good TS values need not be optimal for the calculation of G data. Finally the data for LF, LFC and IACC are displayed in Figure 9c. It shows which of the programs carry out the calculations at all (at the date of the dead-line for phase I in 1997) and at which error rate.
Table V. Given surface data in phase II of Round Robin II.

<table>
<thead>
<tr>
<th>Surface</th>
<th>125 Hz</th>
<th>250 Hz</th>
<th>500 Hz</th>
<th>1000 Hz</th>
<th>2000 Hz</th>
<th>4000 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>wooden floor (stairs and stage)</td>
<td>0.15</td>
<td>0.08</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>wooden walls and undersurface of balcony (birchwood)</td>
<td>0.21</td>
<td>0.12</td>
<td>0.09</td>
<td>0.06</td>
<td>0.09</td>
<td>0.13</td>
</tr>
<tr>
<td>ceiling</td>
<td>0.20</td>
<td>0.15</td>
<td>0.10</td>
<td>0.08</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td>reflectors above the stage</td>
<td>0.12</td>
<td>0.10</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>seats in main area</td>
<td>0.45</td>
<td>0.60</td>
<td>0.73</td>
<td>0.80</td>
<td>0.75</td>
<td>0.64</td>
</tr>
<tr>
<td>seats in rear part</td>
<td>0.50</td>
<td>0.66</td>
<td>0.80</td>
<td>0.88</td>
<td>0.83</td>
<td>0.70</td>
</tr>
<tr>
<td>linoleum floor between seats (balcony)</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>back wall and walls without wood veneer, variable reflectors</td>
<td>0.02</td>
<td>0.06</td>
<td>0.06</td>
<td>0.04</td>
<td>0.07</td>
<td>0.10</td>
</tr>
<tr>
<td>window glass</td>
<td>0.02</td>
<td>0.06</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>plastered concrete (front and undersurface of speaker rooms)</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>ventilation grid on the stage</td>
<td>0.08</td>
<td>0.12</td>
<td>0.15</td>
<td>0.15</td>
<td>0.12</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Diffusivity

| for big flat surfaces                                       | 0.15   | 0.13   | 0.11   | 0.09    | 0.07    | 0.05    |
| for smaller flat surfaces and misc. surfaces               | 0.30   | 0.27   | 0.24   | 0.21    | 0.18    | 0.15    |
| for auditorium surfaces                                    | 0.30   | 0.40   | 0.50   | 0.60    | 0.65    | 0.70    |
| for closed adjustable reflectors                           | 0.15   | 0.20   | 0.25   | 0.35    | 0.45    | 0.50    |

Table VI. Conditions of calculation of the participants on phase II Participants: Commercial software: P2, P5, P9, P16, users of commercial software: P3 = user of P5, P15 = user of P16, P13 user of other commercial software.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Rays/oct/source</th>
<th>No. of surfaces in the model</th>
<th>Computing time</th>
<th>Hardware used</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>18h 00 min</td>
<td>DEC Alfa server 4100</td>
</tr>
<tr>
<td>2</td>
<td>20000</td>
<td>272</td>
<td>2h 20 min</td>
<td>Pentium 300 MHz</td>
</tr>
<tr>
<td>3</td>
<td>2500</td>
<td>147</td>
<td>25 min</td>
<td>Pentium 133 MHz</td>
</tr>
<tr>
<td>5</td>
<td>1504</td>
<td>94</td>
<td>86 s</td>
<td>Pentium II 300 MHz</td>
</tr>
<tr>
<td>DXF</td>
<td>7520</td>
<td>470</td>
<td>116 s</td>
<td>Pentium II 300 MHz</td>
</tr>
<tr>
<td>6</td>
<td>10000</td>
<td>-</td>
<td>4h 06 min</td>
<td>Alfa Station 200</td>
</tr>
<tr>
<td>6DXF</td>
<td>10000</td>
<td>566</td>
<td>4h 43 min</td>
<td>Alfa Station 200</td>
</tr>
<tr>
<td>8</td>
<td>1500000</td>
<td>165</td>
<td>30 h 00 min</td>
<td>Pentium 133 MHz</td>
</tr>
<tr>
<td>9DXF</td>
<td>2048</td>
<td>565</td>
<td>42 s</td>
<td>Pentium Pro 200 MHz</td>
</tr>
<tr>
<td>13DXF</td>
<td>10000</td>
<td>566</td>
<td>20 min</td>
<td>PC 486</td>
</tr>
<tr>
<td>14DXF</td>
<td>41167</td>
<td>566</td>
<td>60 min</td>
<td>Pentium Pro 200 MHz</td>
</tr>
<tr>
<td>15DXF</td>
<td>-</td>
<td>566</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16DXF</td>
<td>50000</td>
<td>566</td>
<td>20 min</td>
<td>IBM RS/6000</td>
</tr>
</tbody>
</table>

3.4.3. Phase II: surface properties prescribed

In the starting phase of the Round Robin, the participants had to estimate the surface properties, therefore their personal experience and skills on acoustic design of rooms played an important part. The results may be regarded as a first guess, leaving many free parameters to the estimation of the participants. The specification of absorption and diffusivity in phase II should lead to the real differences in the programs being revealed. Unfortunately it was not possible to measure the surface data in situ and provide a more reliable basis for the calculations and comparisons with the parameters measured. The given values therefore could only be another guess, and the differences from the measured values even in the case of an ideal error-free program are indicative only of the quality of this estimation of absorption and diffusivity. The absorption data were adopted from the standard literature (e.g. [14]) and adapted to the conditions prevailing in the ELMIA hall, while the diffusivity data had to be estimated. Nevertheless the comparison of the different calculation results will give an estimation of the variance of room simulation software. Since also the number of DXF-model users has increased, the common data input in this phase is improved and the influence of the operator is reduced significantly.

Table V shows the surface data given to the participants in phase II of the Round Robin. Unfortunately not all participants in phase I could perform the calculations for phase II; on the other hand, some new participants joined the Round Robin. In the time between the start of phase I (October 1996) and the end of phase II (July 1998) the software was improved on the basis of further developments, e.g. an extension of the number of calculated parameters by $I_{ACC}$-values. One participant (P9) decided to change his geometry model from a 3530 surface model to the DXF model containing only 566 surfaces in order to reduce the computing time.
Some participants partially modified the DXF model for their calculations: they combined surfaces and thus reduced their number.

Although the participants were asked to fill the conditions of calculation in the result form, not all data are available so far. The data known for phase II are shown in Table VI including the number of rays, number of surfaces, computing time and hardware used. It gives an impression of the big differences in the computing times (from 42 seconds to 30 hours), although similar hardware has been used. The software standard today seems to be based on the Windows 95/98/NT platform running on a Pentium PC, especially for commercial software. The great number of rays for P8 indicates that the software used is based on a statistical approach.

In order to show the reduction of the uncertainty of calculation by the common surface data input in Figure 10, the standard deviations $\sigma$ for the calculated $D_{50}$ values in the six octave bands are given for the two phases. For each position SxRx, the corresponding values can be read out, showing in most cases the maximum values in the lowest octave. It can generally be said that the standard deviation is roughly halved by specifying the absorption and diffusivity data. A similar tendency is found for the other parameters. Since the values given for the unknown surface data are also estimated, the mean error of the calculations related to the measurements in this phase need not decrease in the same way. It can also be seen from the figure that some positions are more critical than others, generally the source position S2 shows a somewhat greater values for $\sigma$ in phase II. The critical receiver position R2 on the balcony shows no significant increase for $\sigma$ in the case of the hidden source S2 and only a slightly greater value for S1 compared to the other positions. This may also be due to the great number of DXF model users which increased from four in phase I to seven in phase II, resulting in a more homogeneous geometry input.

For an overall comparison of the relative errors of all the participants in the following figures, mean values of the 12 source-receiver combinations in the six octave bands will first be presented. The errors as defined above will again be related to the subjective liveness in the 1 kHz octave (see Table I). The values for $T_{50}$ and $EDT$ for the 11 participants of phase II (two participants again furnished two results using also the DXF model besides their own model) are shown in Figure 11a. Only one participant (P15) still calculated much too small $T_{50}$ values even with given absorption data, which confirms the assumption that also a software operating error was the reason for this; note that $EDT$ seems to have correctly been calculated by P15. In contrast, P13 this time furnished
significantly better values. For the $EDT$ the overall mean error values show nearly constant values of about twice the subjective limen.

In Figure 11b the results for $D_{50}$, $C_{80}$, $TS$ and $G$ are shown. Although there are differences between the individual participants, it can be assumed that the mean relative error appears to be what is possible under the given circumstances, i.e., given estimated absorption data, application of the rules of geometrical acoustics and a limited number of surfaces. To complete the display of the parameter calculation in Figure 11c, the remaining three quantities $LF$, $LFC$ and $IACC$ are shown as far as supported by the software used. It becomes clear that all these overall mean errors give only a rough overview of the current accuracy of calculation and that it is hard to say whether there is a winner giving the best results for all parameters. For an analysis of the calculation problems, the results presented in the six octave bands is much more informative.

Figure 12a shows the relative mean error for $T_{30}$ averaged over the 12 positions. As expected, the lowest octave shows the biggest error, followed by the 250 Hz octave. Here also the differences between the participants become much more significant than on the overall average and for the higher octaves some participants have relative errors which lie clearly below 2, some even below 1. The calculation of $EDT$ (Figure 12b) appears to be more critical, and no participant has obtained a mean value below 1. The 125 Hz octave again is the most critical one for most participants. The $D_{50}$ graph (Figure 12c) shows a similar tendency, although for the higher octaves the differences between the participants are smaller.

A mean error of roughly once or twice subjective limen can be read out. The $TS$ curves in Figure 12d show an increased relative error especially for the two lowest octaves.

If the performance quality of room simulation programs in general has to be judged, it will be more meaningful to average all the results of the participants at a given position and display these values in the six octave bands. It will more clearly reveal the problems that certain positions imply, e.g., what will happen if the source is invisible for the receiver (S2R2), what parameters show a significantly greater error and what octave furnishes the most uncertain values. Figure 13 will demonstrate for $T_{30}$, $EDT$, $D_{50}$, $TS$ and $G$ where the weak points of calculation are.

The plots of the $T_{30}$ mean values (in s, Figure 13a) first of all shows, that the low frequency reverberation is underestimated by all participants. This error seems to influence only the late decay, since the $EDT$ error for 125 Hz is substantially smaller (see Figure 13b). The overestimation of the absorption of the surfaces in the given absorption data (e.g., the seat data in Table V) may be the source for the high error rates at 125 Hz. This assumption is confirmed by the fact that with his own estimated data (0.1 for 125 Hz) participant P1 got in phase I lower errors for $T_{30}$ than in phase II (see 3.4.2 and Table V). The error for the higher octaves decreases to quite moderate values at 4 kHz.

While the errors for $T_{30}$ appear to be nearly independent of the position, the $EDT$ (Figure 13b) errors show clear variations which may help the programmer to find some weak points in the software. While in the lowest octave S2R2, S2R5 and S2R6 seem to be critical, in the 4 kHz octave, the very
close position S1R1 is calculated with an error four times that of S1R4. As mentioned above, $D_{S0}$ for S1R1 is calculated by all participants with a great error in the lowest octave, but also in the other positions of Figure 13c, the problems in the lowest octave are obvious. Even for the high frequencies S2R2 is critical because of the missing direct sound from S2. This tendency is also found in the graph for the $TS$-error (Figure 13d). The errors for $G$ which has not been calculated by all participants (see Figure 13e) are concentrated on other positions, e.g., R6 appears to be critical for all frequencies and both source positions.

Of course, not all results can be shown and discussed here, therefore it is intended to give access to the complete set of data of the Round Robin to everybody on the website of the PTB (http://www.ptb.de/english/1/14/14011401.html), including the geometry DXF file and the measurements. This would enable any software to be tested by using the ELMIA hall data and comparing them with measurement data. Maybe the data for other rooms may in the future be added, e.g., those of the PTB lecture hall which was the subject of the 1st Round Robin.

4. Conclusion

The results of the 2nd Round Robin that are presented here give an impression of the accuracy of calculation for the present room simulation software. It should be kept in mind that the quality of a computer simulation is strongly dependent on the input data and the feeling of the operator for the absorption data of the surfaces. It can be concluded that a good program cannot replace the skill of a good acoustician, but also the usability plays an important part for the rating of the quality of room simulation software. As expected, the weak points of the programs are concentrated on the low frequency calculations as diffraction effects have been neglected in the geometrical calculations. The error found in the calculations on $D_{S0}$ in the lowest octave (cf. Figure 7 and 8) should be investigated thoroughly by the program development teams in order to detect the source of this systematical error. Further, the influence of sound propagation over seats should be implemented in room simulation software [12]. For a more precise calculation of frequencies below 100 Hz other methods should be applied in the software [8].
The most important prerequisite for a successful room simulation is a precise knowledge about the surface material, i.e., its absorption and diffusivity properties. The great diversity of the estimated data for the seat area (Table IV) indicates that more detailed tables are required for reliable room acoustical calculations. These data should be based on in-situ measurements for absorption as well as diffusivity as proposed by various authors [15, 16, 17]. Most valuable data on occupied and unoccupied seats in concert halls are given in [18]. As shown above, the comparison of calculated values with measured data is also strongly depending on the uncertainties of measurements and of the evaluation procedure applied to the measured impulse responses. Therefore, reference measurements require a verification, especially in the case of LFC/LFLC, sound level (G) and IACC.

During the performance of this Round Robin, it was confirmed by the participants that such a comparison with measured data and with results of other programs is also highly appreciated by the software developers and users as a feedback for their further work. Therefore it is intended to continue this work of international coordination in another Round Robin. In order to get more detailed results, it should be tried to supply the users with measured data of absorption and possibly also diffusivity for this new room. Further, the comparison of calculated auralizations [19] with measured sounds recorded via a dummy head will be an interesting supplement.

Acknowledgement

The author would like to thank all persons who helped to make the 2nd Round Robin a successful project: first of all Clenn Junefelt for the help in collecting all the material about the ELMIA hall and for the coordination of the measurements in Jönköping and also Prof. Mendel Kleiner for the good cooperation and help. Many thanks also to the measuring teams from ITA Aachen, Chalmers University and PTB who had a lot of work when evaluating the data from the measured impulse responses. Further to the IRCAM team (Olivier Wursch and Frederico Cruz-Barney) for verifying the PTB measurement data with their own evaluation software. Thanks also to L Cairovi for opening his DXF model to all participants on the web page of Helsinki University (HUT), to Prof. Y. Ando and Shin-ichi Sato from Kobe University (Japan) for the correspondence about JND on IACC. Last but not least the author wishes to thank all participants for the good cooperation and the many recommendations during the performance of the Round Robin.

References