In modern room acoustical simulation programs, the use of auralisation is one of the most convincing arguments for the demonstration of acoustical changes. The calculated sound files include the acoustical properties of the room transfer path between sender and receiver, which means that the room geometry, absorption, scattering and also diffraction can be taken into account. For a comparison of the auralisation quality of room simulation programs it must be decided whether comparing the calculated binaural impulse responses is sufficient because it includes all relevant information, or whether listening tests with convolved anechoic sound material as input have to be performed instead. The most interesting case, the comparisons with real sound, requires a high degree of adjustment to all sound-influencing parameters within the simulation programs: radiation properties of the sound source(s) (directivity, frequency response), the virtual listener's head related transfer functions (HRTF) and also the reproduction system (headphones, loudspeakers in a transaural system) have to be equalised properly.

1 Introduction

The computational properties of room simulation programs have been compared in three international round robins. In these comparisons, room models had to be built, and as results, the room acoustical parameters as defined in ISO 3382 had to be calculated for a number of source-receiver combinations. These data reflect only one part of the facilities of room simulation programs: the numerical evaluation of these parameters which are derived from the calculated impulse response. Although these parameters tell the skilled acoustician a lot about the room acoustical properties and their frequency dependence, another feature which today is implemented in all current software versions gives a more realistic view into the expected acoustics of a future hall: the so-called "auralisation". By convolving arbitrary anechoic input signals with the impulse responses of the corresponding transfer paths from the sender to the receiver, the programs can produce a realistic binaural sound signal which supplies also for the non-acoustician a virtual three-dimensional acoustic image of the hall under test and of the soundfield inside it excited by the chosen input signal.

For comparing the auralisation properties of room simulation programs, a simple comparison of numerical data as in the case of the room acoustical parameters is not possible. Therefore it has to be checked how the quality of auralisation is influenced by the calculation processes in the software, and which data are suitable for comparing different programs.

2 Auralisation in room acoustical simulation programs

The impulse response for a given source-receiver path according to the usual measuring practice contains the sound energy as collected by an omnidirectional microphone and contains no directional information, although the program 'knows' from which direction each individual sound ray is arriving, at least in the early part of the impulse response. This directional information makes it possible to apply the behaviour of a virtual dummy head: the so-called Head Related Transfer Function (HRTF). HRTFs are today documented for a number of dummy heads. They describe the complex transfer functions of the sound coming from a given direction to the ear canal entrance of both ears of the dummy. These data contain frequency-dependent shadowing effects and diffraction of the head and the ears, and in most cases (dummy heads with torso) also reflections from the shoulders. These HRTFs are applied to the monaural impulse responses and result in stereophonic (binaural) impulse responses which can be used for convolution with anechoic sounds. If the resulting sounds are played via headphones, a more realistic three-dimensional image of the room under test will be presented to the listener. This includes spatial impressions of rooms, e.g. the direction of discrete reflections (echoes), a feeling for the room size, the apparent source width etc. [1].
3 Differences in room simulation programs relating to auralisation

The general procedure for transformation of omnidirectional to binaural impulse responses by using the HRTFs is similar in all programs, but it may differ in some details:

1. The dummy model from which the HRTFs are used can be different.
2. The spatial resolution and distribution of measured incidence directions, e.g. equal angles in all directions or concentration in some areas (front).
3. Frequency response equalisation of the virtual dummy head: free field - diffuse field (included in the HRTFs)
4. Resolution of frequency responses of HRTFs (1/3 oct, 1/1 oct, ..)
5. Frequency response of the headphones-equalisation in the programs

These differences will lead not only to different sound colourings but also to differences in the spatial impression of sounds at headphone reproduction.

4 Considerations with regard to comparison problems

The comparison of the auralisation properties of different programs requires a minimum of uncertainties for the data input. The experience with a very simple 7-planes model in Round Robin III has shown that even in this case the calculations of inexperienced users may lead to wrong results [2]. A better solution could therefore be, to carry out such a comparison in cooperation with the software-developers under well-defined conditions.

4.1 Defining a common base for data input

Four main aspects will influence the quality of the result of a comparison using identical input data:

1. The calculation procedure of the omnidirectional impulse response of the transfer path,
2. The application of the HRTFs for the calculation of the binaural impulse response,
3. The reproduction system used for binaural listening (headphones, wave field synthesis, transaural system, head tracker),
4. Specification details of non-omnidirectional sound sources (frequency-dependent directivity)

Furthermore, the convolution software which is included in the simulation software package can influence the quality, but it is assumed that this mathematically well-defined procedure, which uses the overlap-add algorithm for convolving long source signals, should not introduce software-specific errors.

Consequently, it would be possible to compare only the binaural impulse responses of the different programs first and investigate the differences and to compare the results then with a measured binaural impulse response of the real room [3]. In a second step, convolution with special unechoic sound material could be performed by using a single convolution program. This would make it possible to listen to special sound material for judging auralisation quality in subjective tests.

4.2 Sound source properties

For the auralisation of realistic sound sources (musical instruments, voice, loudspeakers, machine noise), the sound source properties have to be different from the omnidirectional source as required for standard measurements of room acoustical parameters according to ISO3382. The application of the source properties (frequency-dependent directivity) in the various programs may be different and a proper definition of the conditions is needed. Similar to the realisation of recording of incoming sound rays from different angles in space (HRTF), the source signal needs a proper specification if it is not omnidirectional. Depending on the spatial resolution of the applied source model and on the distribution of sound data in frequency bands of a certain size, even the sound data input for the programs is difficult to define unequivocally.

4.3 Comparison with real sounds

The most interesting but also most demanding case for an auralisation is the comparison with real sounds. Here some more requirements have to be fulfilled concerning recording and/or presentation of the real reference sound. A well-defined direction of spatial orientation, knowledge of the recording position of the real reference sounds and a proper level adjustment are required.

The case of a real (living) reference, e.g. a speaker talking to the listeners, in contrast to a simulation played via headphone or a synthetic sound field (wave field synthesis or transaural system with head tracker), will be very hard to realise and reproducibility is one of its main problems. But as experience with auralised sound of human speakers has shown, there are also some non-linear effects audible, especially at the reproduction of critical consonants (sibilants) which
enable a clear subjective distinction between simulation and real sound.

5 First experiments

For getting an impression of the differences in the HRTFs applied in current software, the following situation was modelled in four different room simulation programs (CATT, EASE, Odeon, Ramsete):

In a virtual room with totally absorbent surfaces \((a = 100 \%)\), pink noise and male speech were radiated by an omnidirectional source at 1.5 m height. A dummy head receiver which was supplied by the software at a distance of 3.5 m was directed towards the source at a height of 1.2 m (according to ISO3382 recommendation). The binaural impulse response of all four programs was convolved with the two sounds. The results for the stationary pink noise could be used for measuring the spectral differences; they are displayed in the Figure 1. This signal is also optimal for subjectively testing the sound coloration in the four cases in comparison with the source signal as reference via headphones. Since the frontal position of the source relative to the dummy head has equal HRTFs to both ears, also the symmetry properties could be compared.

The subjective evaluation confirmed the differences in the measured pink noise curves and they were clearly audible. Also for the male speaker, severe colorations were noticed immediately after switching from one example to the next, although no spatial differences were audible.

In another experiment the simple seven walls model of Round Robin III phase 1 [2] was used to compare also the spatial image produced by the heads. In this case not only the HRTFs are responsible for the sound but also the room simulation process. Listening via headphones, apart from the sound coloration also substantial differences could be registered concerning the sound coming from the sides and the apparent source width.

![Figure 1: Spectral reproduction of pink noise under virtual unechoic conditions using four different room simulation programs; simulation of binaural recording at a distance of 3.5 m](image-url)
6 Summary

General considerations and preliminary experiments confirm that a comparison of the auralisation properties of room simulation programs depends not only on the calculation of the monaural impulse response of a transfer path in virtual room model but also, to a high degree, on the applied HRTFs of the virtual dummy head. A round robin for comparing the quality of auralisation needs a common base for the application of HRTFs for all participants in order to get the same conditions for evaluating the simulation of spatial quantities. It seems necessary to apply listening tests for the subjective aspects of the room image and the quality of speech reproduction.

References

