Auralization of Impulse Responses Modeled on the Basis of Ray-Tracing Results 3122 (C-4)

K. H. Kuttruff
Technische Hochschule Aachen
Aachen D-5100, Germany

4E/C-4

Presented at the 91st Convention 1991 October 4–8 New York

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

AURALIZATION OF IMPULSE RESPONSES MODELLED ON THE BASIS OF RAY-TRACING RESULTS

K.H.Kuttruff

Technische Hochschule Aachen, Templergraben 55, D5100 Aachen, Germany

ABSTRACT

After a short comparison of the most common methods of acoustical room simulation - ray-tracing and image source model - it is shown that the information contained in a ray-tracing histogram is sufficient for the auralization process. Furthermore, the steps are discussed which are needed to transform this information into a binaural impulse response ready for the convolution with music signals.

INTRODUCTION

The term "auralization" is to denote a new procedure which enables us to create aural impressions from computer simulated room responses. The simulation itself can be based upon acoustical and geometrical data of existing rooms as well as of non-existing ones. Thus, auralization can be employed to present music, performed in halls which are still under design, or even in completely hypothetical halls which no architect would ever take into consideration. Furthermore, it is going to become an important tool in room acoustical research since it allows us to study the influence of size, shape, materials, arrangement of audience seats and many other factors on the acoustics of a hall.

The idea of auralization is quite old. As early an in 1934, Spandöck [1] proposed to replay frequency-transformed music signals in a scale model, to pick them up again in the model and to present them by earphone to a listener after re-transformation into the original frequency range. A similar procedure employing computer models was proposed in 1961 by M.R. Schroeder and his co-workers [2] and was first applied to simple rectangular enclosures by Pösselt [3]. Since them, several authors have reported on auralization schemes and the results they obtained with them [4, 5, 6].

Basically, auralization consists in the construction or simulation of a digital filter which reproduces both the transmission of sound signals in the considered room and their binaural reception by the listener. The latter requirement is accounted for by including the listener's personal head transfer functions into the filter characteristics.

The present paper concentrates on the way in which the room response is simulated and how the results of this simulation is transformed into the characteristics of the auralization filter.

SIMULATION OF ROOM RESPONSE

To simulate the physical transmission of sound signals in an enclosure, two different procedures are currently employed, namely ray-tracing and the image source method.

In a ray-tracing procedure as first applied by Krokstad, Strøm and Sørsdal [7] to concert hall acoustics, a sound source is imagined to release numerous sound particles into all directions at a certain moment. The path of each particle - including its reflections from specularly or diffusely reflecting walls - is followed up with the computer, and whenever it arrives at a previously assigned counter, its energy, arrival time, direction etc. is registered. The result of this process can be presented as a histogram (see Fig. 1), showing the temporal distribution of the recieved energy, and can be regarded as an approximation to the energetic impulse response of the enclosure with the degree of approximation depending on the number of particles and the achieved time resolution.

The image source method is based on the old idea that a sound ray which is reflected from a plane wall can be thought of as originating from an image source which is the mirror image of the original sound source formed by the wall. Accordingly the construction of individual ray paths is replaced with the construction of image sources not only of first order, but also of second, third etc. order (images of image sources). Once all significant image sources have been found the resulting signal in a certain room point is obtained just by adding all their contributions. If the sound signal is a Dirac impulse, the result of this process is the correct impulse response of the enclosure. An example, calculated for frequency-independent reflectivities of all walls, is shown in Fig. 2.

To obtain realistic impulse responses, however, the frequency dependence of wall reflectivities (and air attenuation) must be accounted for in some way, for instance by convolving each spike in Fig. 2 with the "reflection responses" $r_n(t)$ of all walls which are involved in the formation of that particular component. Here, $r_n(t)$ is the Fourier transform of a wall's complex reflection factor $R_n(\omega)$.

At first glance, ray-tracing results as shown in Fig. 1 are less suitable as a basis for the digital filter we need for auralization. On the other hand, in the image source model the number of significant images sources grows exponentially with the desired length of the impulse response. Moreover, only a small fraction of all image sources is "visible" from a given receiving point. Although efficient schemes have been developed to detect only the visible images and to omit the invisible ones [8, 9], this method still requires long computing time. Furthermore it yields much more information on the physical transmission characteristics of the room than is significant from a psychoacoustical point of view: Due to its limited temporal and directional resolution, our hearing is unable to preceive all details of the physical impulse response. This holds for the directional distribution of reflections, as well as for their arrival times and strengths.

MODELLING THE IMPULSE RESPONSE

For these reasons, it is both desirable and possible, to reduce the information contained in a physical room impulse response and to replace it by an "equivalent response" which is much simpler. It is essential, of course, that an auralization filter with this "equivalent response" creates exactly the same subjective impression as the original response, i.e. that music passed through this filter sounds the same way as it would sound to a listener if he were sitting in the considered hall.

For the construction of this equivalent response one should keep in mind, that two portions of a room's impulse response with different subjective effects can be distinguished. The first one consists of the direct sound and the earliest reflections, its temporal and directional structure is of crucial importance for the sensation of loudness, clarity, spaciousness etc [10]. Furthermore, it is responsible for the variation of listening conditions within different parts of a hall.

The remaining part, the "late portion" of the impulse response consists of numerous overlapping reflections. This is what is usually known as reverberation. Although also important for the hearing impression as far as the decay rate and its spectral changes are concerned, it is less specific for the acoustics of a hall and its various parts.

Accordingly, the equivalent impulse response can be built up of two parts (Fig. 3), the first consisting of correctly simulated reflections determined from image sources or with high-resolution ray-tracing (interval width 20 μs). In the formercase, the image sources are derived by a simple back-tracing process from ray-tracing data [11]. The "late portion" is synthesised in such a way, that its spectral composition is the same as that of the original reverberation, but without the complicated "microscopic" structure of the latter. It can be derived from a relatively coarse simulation procedure, namely a ray-tracing procedure with low time resolution, simular to that presented in Fig. 1.

The characteristic delay time t_C which marks the end of the early and the beginning of the late portion can be determined by regarding the directional distribution of reflected sounds. This distribution is quite irregular immediately after the direct sound has arrived, but with increasing time delays of the reflections it becomes increasingly uniform, due to the randomizing effect of the enclosure. After 100 - 150 ms the sound field can be regarded as diffuse [11] and does not contain any directional information. Thus it is reasonable to set $t_c \approx 150$ ms.

The late portion of the equivalent impulse response can be modelled in the following way:
The whole frequency range is subdivided into a number, say 10, frequency bands, octave bands, for instance. For each of them, a ray-tracing simulation is carried out with the proper wall absorption coefficient and air attennuation, while the geometrical data (room shape, location of sound source and receiver etc) remain unaltered. A time resolution of 5 or 10 ms seems to be sufficient. The results can be represented by numbers E_{ik} (Fig.4) with indices i and k denoting the number of the time interval and of the frequency band, respectively. Thus the set of values with constant i represents the short-time energy spectrum of reverberation in the time interval Δt_i. - The next step is to find a time function $g_i(t)$ which has the same energy spectrum as the reverberation. It is obtained from the E_{ik} (constant i) by square-rooting, interpolation and Fourier transformation the modified spectrum into the time domaine. The latter step requires knowledge of the phase spectrum, strictly speaking, which is not contained in the E_{ik}. - Now Poisson-distributed points t_n on the time axis are selected on the time axis by properly distribute random numbers. Then the equivalent response within the time interval Δt_i is obtained by

$$g(t) = \sum_n g_i(t - t_n) \qquad (1)$$

The density of the t_n on the time axis should be in the range of 1000/s.

Because of the random superposition of the elements $g_i(t)$ according to Eq.(1) the phase spectrum used for the above-mentioned Fourier transformation is uncritical and can be chosen in any reasonable way which is convenient from the computational point of view. In fact, listening tests with various phase spectra including one corresponding to a minimum phase system have shown that the differences cannot be noticed.

BINAURAL PROCESSING

As mentioned in the introduction, binaural hearing has to be included in an auralization filter to convey natural impressions on the spatial structure of the sound field to a listener. Usually this is achieved by incorporating his head transfer functions into the characteristics of the final filter, which results in two final filters, one for each ear. These transfer functions depend on the direction from which a particular reflection arrives at the listener's head. Due to the limited directional resolution of our hearing it is certainly sufficient to group all directions into a number of "directionals classes". The minimum number of these classes and their sizes and boundaries are still under investigation.

For the late portion of the impulse response the binaural processing is particularly simple: Since for $t > tc$ the sound field is diffuse, we can allocate the $g_i(t - t_n)$ in Eq.(1) randomly to directional groups, but in such a way, that also in the equivalent response all directions are uniformly distributed.

RESULTS AND LISTENING TESTS

Auralzation filters designed after the described scheme have been programmed in various modifications by R. Heinz and R. Uekermann. The enclosure they simulated was an existing concert hall with a volume of about 15 000 m^3. Therefore, it is possible to compare music samples processed with the auralization filter on the one hand with those obtained by convolving the samples with the real impulse response as recorded binaurally for the same situation in the real hall on the other.

It turned out that the width of time intervals Δt_i had to be chosen as small as 1 ms. But still the required computing time was within reasonable limits. Special care was needed for joining both parts of the equivalent impulse response.

Subjective comparisions of the music samples showed almost perfect agreement, samples processed with the auralization filter and the natural impulse response are nearly undistinguishable. For reasons which are unknown so far the agreement was somewhat less perfect for speech samples but it was still very good.

CONCLUSION

The modelling scheme described in this paper has considerable advantages over the image source method as usually employed. In the first place, it requires much less computing time than the latter. Furthermore, no knowlege of the complex wall reflection factors is needed, which are difficult to collect. The same holds for the involved loss processes of air. Instead it is suffficient to base the design of the auralization filter on the absorption coefficients and air attenuation constants which can be looked up in tables.

Nevertheless, several details of the procedure need closer investigation, and further improvements and simplifications are certainly possible. But even in its present state, these method seems to fullfill the expectations.

The author is indebted to the Deutsche Forschungsgemeinschaft which supports this project by a financial grant.

REFERENCES

[1] F. Spandöck
Ann. d. Physik V., 20 (1934), 345

[2] M.R. Schroeder, B.S. Atal, C. Bird
Proc. 4th Intern. Congr.(Acoustics, Copenhagen (1962), paper M 21

[3] C. Pösselt
Fortschr. d. Akustik DAGA '87, DPG-GmbH, Bad Honnef (1987), 725

[4] T. Kihlman
Proc. 21th Conf. Acoustics, Strbské Pleso, Czechoslovakia, Vol 1, 104

[5] J. Martin, J.P. Vian
Proc. 13th Intern. Congr.Acoustics Belgrade, (1989), Vol. 2, 253

[6] H. Kuttruff, M. Vorländer, Th. Classen
Acustica 70 (1990), 230

[7] A. Krokstad, S. Strøm, S. Sørsdal
J. Sound Vibr. 8 (1968), 118

[8] P. Vian, D. Van Maercke
Proc. Vancouver Symp. (12th ICA Congr. 1986), 74

[9] M. Vorländer
J. acoust. Soc.Amer. 86 (1989), 172

[10] L. Cremer, H.A. Müller
Principles and applications of room acoustics, Vol. 1
Applied Science Publishers, London 1982

[11] J.W. Junius
Acustica 9 (1959), 289

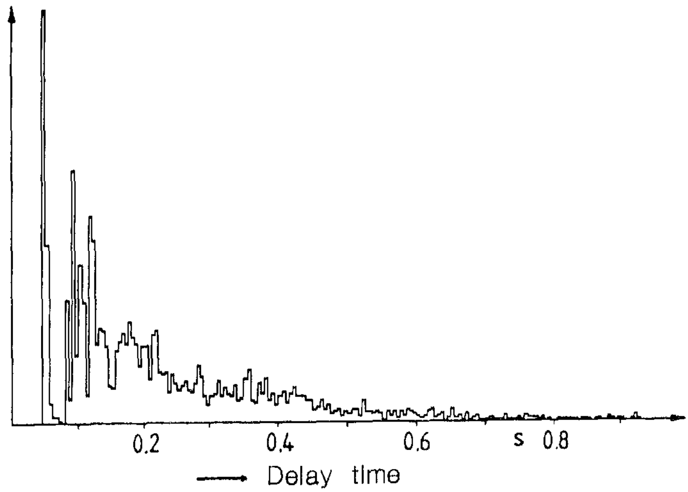

Fig. 1: Typical ray-tracing histogram. The interval width is 5 ms

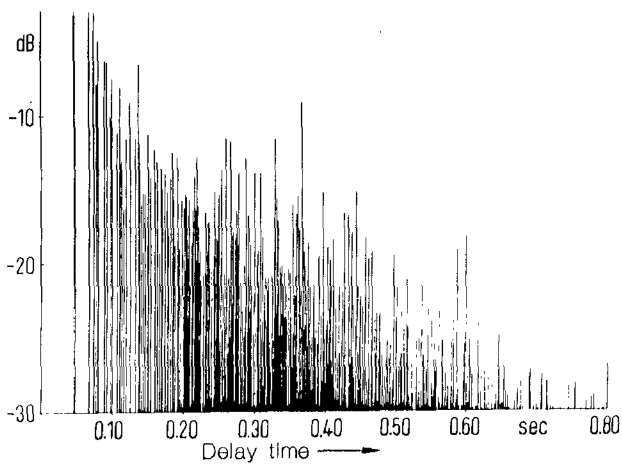

Fig. 2: Energetic impulse response obtained with the image source method

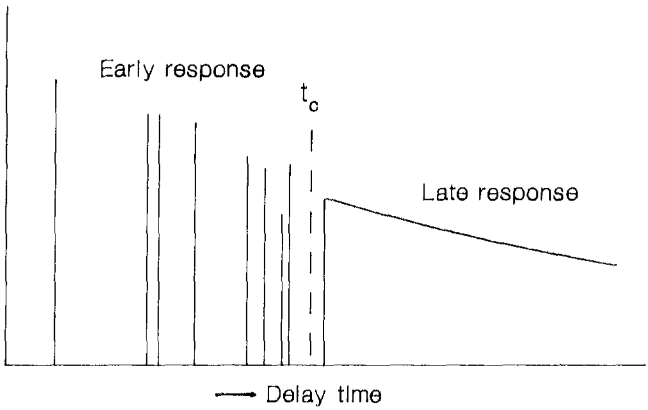

Fig. 3: Early and late portion of an impulse response

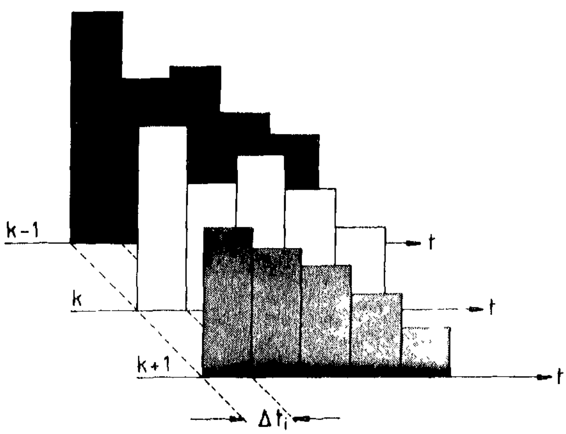

Fig.4: Modelling the "equivalent impulse response"
(i = time index, k = band number)