Room in Room Acoustics: Using Convolutions to find the Impact of a Listening Room on Recording Acoustics

C.C.J.M. Hak (c.c.j.m.hak@tue.nl)  
R.H.C. Wenmaekers (r.h.c.wenmaekers@tue.nl)

Department of the Built Environment  
Building Physics and Services  
Eindhoven University of Technology  
P.O. Box 513,  
5600 MB Eindhoven, The Netherlands

ABSTRACT

From experience and earlier investigations it is clear that room acoustical details in recorded music or speech can only be heard in a room having a reverberation time shorter than the one in the room in which the recording was made. The acoustical properties of listening rooms influence the perceived acoustics of the recording. In earlier investigations, the practical impact of listening room impulse responses on recording room impulse responses was shown by convolving many random combinations of measured room impulse responses. For more insight in the impact of listening room acoustics on rendered acoustics, in this new research, convolutions of synthetic impulse responses have also been used. Both the effect of the decay rate and the amount of direct sound were taken into account, where the resulting change in an acoustical property is assumed to be negligible if it does not exceed the JND (Just Noticeable Difference). Both theoretical and practical cases show that during playback, the decay curve derived from the recorded impulse response turns into a curve with a slow attack and a concave level decay line. The more both impulse responses (recording and playback room) are diffuse and equal in decay rate, the higher the impact. Even when using nearfield playback, it is very difficult to reduce the negative impact of a listening room on acoustical details in a recording.

1 INTRODUCTION

From experience and an earlier investigation\(^1\) it is clear that a recorded reverberation time can only be demonstrated in the diffuse field of a room having a reverberation time shorter than the one in which the recording was made. The smallest details and the finest nuances with regard to colouring, definition and stereo image can only be judged and criticized when there is little acoustical influence from the playback acoustics on the recorded acoustics. However, usually the listening or playback room in combination with the used sound system affects the recorded acoustics. This happens in class rooms, congress halls, cinemas and even in sound control rooms. In this research, theoretical convolution results have been compared with experimental values. Both the effect of the decay rate and the amount of direct sound in the listening room were taken into account.
2 THEORETICAL BACKGROUND

The convolution $y$ of signal $s$ and system impulse response $h$ is defined and written as:

$$y(t) = s(t) \ast h(t)$$

(1)

or

$$y(t) = (s \ast h)(t) = \int_{-\infty}^{\infty} s(t) \cdot h(t - \tau) d\tau$$

(2)

From a room acoustical point of view $s(t)$ is a sound that is recorded in an anechoic room (dry recording) and played back in a standard room, $h(t)$ the impulse response of the standard, more or less reverberant room and $y(t)$ the convolved sound as it is heard in that standard room. Therefore, an impulse, for instance a hand clap, recorded in an anechoic room, played back in a reverberant room, is heard as an impulse response of that reverberant room. A recorded impulse in the reverberant room that is played back in the anechoic room is again heard as the impulse response of the reverberant room (associativity in convolution algebra). In both cases the derived room acoustic parameter values will be the same.

When both the recording room and the playback (listening) room are reverberant, smoothing of the sound occurs. Therefore, in some cases it is impossible to judge the original recordings in detail. The room acoustics in the sound recording that we want to demonstrate or judge will be affected by the acoustics of the listening room. With a double convolution by which an impulse response from one room is convolved with a dry recording and afterwards the result is convolved with the impulse response of another room, it is possible to hear how a recording, made in a reverberant room, sounds when played in another reverberant room. The result is usually an unwanted smoothed sound signal. By using a pure impulse (Dirac delta function) instead of a normal sound signal to be convolved with both room impulse responses (eq 3 and 4) we can examine what one room does with the other concerning the values for the room acoustic parameters (eq 5). So it is possible to derive a ‘room in room’ acoustic parameter value from the smoothed room impulse response RIR or its derived energy time curve ETC (Figure 1).

Figure 1: Example of the effect of ‘room in room’ acoustics on measured impulse responses and level decay curves.
Mathematically:

\[ h(t) * \delta(t) = \delta(t) * h(t) = h(t) \]  \hspace{1cm} (3)

where:

\( h(t) \) = room impulse response
\( \delta(t) \) = Dirac delta function (ideal impulse)

\[ h_{12}(t) = \delta(t) * h_1(t) * h_2(t) = h_1(t) * h_2(t) \]  \hspace{1cm} (4)

where:

\( h_{12}(t) \) = ‘total’ impulse response room 1 * room 2
\( h_1(t) \) = impulse response room 1
\( h_2(t) \) = impulse response room 2

Substituting equation (4) into equation (1) results in:

\[ y_{12}(t) = s(t) * h_{12}(t) \]  \hspace{1cm} (5)

where:

\( y_{12}(t) \) = convolution of a random sound signal with the ‘total’ impulse response
\( s(t) \) = any sound signal

Starting with the convolution definition:

\[ h_{12}(t) = (h_1 * h_2)(t) = \int_{u=0}^{t} h_1(u) \cdot h_2(t-u) du \]  \hspace{1cm} (6)

and two exponential functions:

\[ h_1(t) = e^{x_1t} \quad \text{and} \quad h_2(t) = e^{x_2t} \]  \hspace{1cm} (7)

the convolution result can be written as:

\[ h_{12}(t) = e^{x_1t} * e^{x_2t} = \begin{cases} 
\frac{e^{x_1t} - e^{x_2t}}{x_1 - x_2} & \text{if } x_1 \neq x_2 \\
\frac{t e^{x_1t}}{x_1} & \text{if } x_1 = x_2
\end{cases} \]  \hspace{1cm} (8, 9)

Starting from a theoretical diffuse sound field, the exponential decay can be written as:

\[ I(t) = I_0 e^{-2at/T} \]  \hspace{1cm} (10)

where:

\( I(t) \) = sound intensity at time t [W/m\(^2\)]
\( I_0 \) = sound intensity at t = 0 [W/m\(^2\)]
\( a = 3\ln 10 \)
\( T \) = Reverberation time defined as the time to drop 60 dB after turning off the sound source [s]
By substituting \(-a/T_1\) for \(x_1\) and \(-a/T_2\) for \(x_2\) the expression \(h_{12}(t)\) for the total impulse response of room 1 * room 2 is obtained:

\[
h_{12}(t) = e^{-at/T_1} * e^{-at/T_2}
\]

where:
\(T_1 =\) Reverberation time of room 1
\(T_2 =\) Reverberation time of room 2

3 ROOM ACOUSTIC PARAMETERS

Many room acoustic parameters are derived from the room’s impulse responses according to ISO 3382-1 \(^2\) and IEC 60268-16. \(^3\) Examples of such parameters are the reverberation time, which is related to the energy decay rate, the clarity, the definition and the centre time, which are related to early to late energy ratios, the speech intelligibility, which is related to the energy modulation transfer characteristics of the impulse response and the laterale energy fraction, the late lateral sound energy and the inter-aural cross correlation, which are derived from lateral impulse responses. In this research two of them have been used, being the reverberation time \(T_{20}\), and the centre time \(t_c\).

3.1 Reverberation time \(T\)

The reverberation time \(T\) is calculated from the squared impulse response by backwards integration \(^4\) through the following relation:

\[
L(t) = 10 \log_{10} \left( \frac{\int_0^\infty p^2(t)dt}{\int_0^\infty p^2(t)dt} \right) \quad [dB]
\]

where \(L(t)\) is the equivalent of the logarithmic decay of the squared pressure. For this investigation the \(T_{20}\) with its evaluation decay range from -5 dB to -35 dB is used to determine \(T\). The just noticeable difference for \(T\) is 5 to 10\%. \(^5,6\) With this parameter it is easy to compare results numerically and graphically (decay line slope).

3.2 Centre time \(T_c\)

The parameter \(T_c\) \(^7\), which is the time of the centre of gravity of the squared impulse response, can be expressed in ms using the following relation:

\[
T_c = \frac{\int_0^\infty t \cdot p^2(t)dt}{\int_0^\infty p^2(t)dt} \cdot 1000 \quad [ms]
\]

\(T_c\) is a room acoustic parameter related to the perceived definition or the balance between clarity and reverberance and avoids the discrete division of the impulse response into early and late
reflections or energy like with for instance the clarity C80.\textsuperscript{8} The just noticeable difference for $T_S$ is 10 ms.\textsuperscript{2} Also with this parameter it is easy to compare results numerically and graphically (shift of maximum graph value).

4 THEORETICAL RESULTS

4.1 Diffuse field in diffuse field (analytical)

Figure 2 shows the theoretical convolution of purely exponential impulse responses. The left graph presents squared impulse ($\sim$ intensity) responses and the right graph the derived decay curves as a result of convolving. An exponential reference impulse response (the recording room, $T$) has been convolved with itself and four other exponential impulse responses (listening rooms). Note that the decay curve derived from the recorded impulse response ($T \text{ conv } T$) turns into a curve with a slow attack and a slight concave level decay line.

![Figure 2: Convolution of purely exponential impulse responses. Left: squared impulse responses ($I_{rel}(t)$), right: decay lines ($L_{ref}(t)$)](image)

4.2 Diffuse field in direct field (numerical)

Figure 3 shows a schematic view of an Energy Time Curve in a predominantly direct sound field. The high initial peak can be described as the difference between the PNR (Peak-to-Noise Ratio) and the INR (Impulse Response to Noise Ratio).\textsuperscript{8}

![Figure 3: Schematic Energy Time Curve (ETC) derived from a room impulse response. The Peak to Noise Ratio (PNR) and the Impulse response to Noise Ratio (INR) can be used to describe the initial peak level: Initial peak level = PNR − INR.\textsuperscript{9}](image)
Figure 4 illustrates the theoretical effect of the direct sound part (the initial peak) of the listening room decay curve on the ‘recorded’ (exponential) impulse response. The left graph shows the effect on the ‘recorded’ reverberation time (in this case $T_{\text{ref}} = 1\,\text{s}$); the right graph shows the effect on the centre time. For both parameters the initial peak has been increased from 5 to 100 dB. Following the graphs it is shown that in order to avoid the negative effect of listening room acoustics on played back recorded acoustics (i.e. without exceeding the Just Noticeable Difference) in some cases an almost impractically high level of direct sound is needed.

![Graphs showing theoretical effect of direct sound part](image)

**Figure 4:** The theoretical effect of the direct sound part (the initial peak) of the ‘listening room’ decay curve on the played back ‘recorded’ (exponential) impulse response. Left: the effect on the reverberation time, right: the effect on the centre time.

## 5 MEASUREMENT RESULTS

For all direct and diffuse field impulse response measurements the same equipment has been used, consisting of the following components:

- **software:** DIRAC (B&K/Acoustics Engineering - Type 7841)
- **input/output:** USB audio device (Acoustics Engineering - Triton);
- **power amplifier:** (Acoustics Engineering - Amphion);
- **sound source:** omnidirectional (B&K - Type 4292);
- **microphone:** ½” omnidirectional, sound level meter (Rion - NL21);
- **microphone:** ½” omnidirectional ICP (Bruël & Kjaer Type 4189);
- **signal:** MLS or eSweep, synchronous or asynchronous

### 5.1 Diffuse field in diffuse field convolution

From earlier research a set of 11 impulse responses has been used. All impulse responses have been obtained from diffuse sound field measurements based on deconvolution techniques using MLS and e-sweeps, resulting in decay range (INR) values > 50 dB. Figure 5 shows the results of the convolutions. The graph shows the difference between 2 values of the reverberation time $T_{20}$, one calculated from $h_2$ and one from $h_1$. On the x-axis the ratio $T_{20}(h_1)/T_{20}(h_2)$ of the reverberation time calculated from $h_1$ ($= T_{\text{recorded room}}$), and the reverberation time calculated from $h_2$ ($= T_{\text{listening room}}$) is given. For symmetry reasons ($h_2 = h_1^* h_2 = h_2^* h_1$), only impulse response pairs with $T_{20}(h_1) > T_{20}(h_2)$ have been depicted. The differences are presented as errors from the real reverberation time $T_{20}$ of the ‘recording rooms’.
Figure 5: Differences between $T_{20}$ of a ‘recording room’ and the $T_{20}$ of a ‘playback room convolved with the same recording room’, presented as a relative error. The solid line is calculated from theoretical purely exponential impulse responses.

The solid line indicates the theoretical errors calculated from purely exponential impulse responses. As can be seen, the first part of the line coincides with the upper limit measurement values and the tail of the graph follows the trend line of the measurement data.

5.2 Diffuse field in direct field convolution

The first measurement results with regard to direct field convolution are presented in figure 6. Impulse responses were measured in a sound control room ($T_{avg}$ approx. 0.2 s) at six different distances from a so called near-field loudspeaker: 5, 10, 20, 40, 80 and 160 cm. These responses were convolved with impulse responses measured in a room with more or less the same reverberation time. Although these first results are obtained only from a few measurements, the effect is clear. Increasing the direct/diffuse ratio by decreasing the listening distance results in a better reproduction of the recorded acoustics. However, for more details further research is needed.

Figure 6: The positive effect of direct sound on sound reproduction. Reduction of the listening distance to a loudspeaker results in an approach of the real (measured or recorded) room acoustical parameter value.
6 CONCLUSIONS

- The theoretical parameter value errors for $T$ calculated from purely exponential impulse responses are in agreement with the errors obtained from practical measurements. Up to a $T_{\text{recording room}}/T_{\text{listening room}}$ ratio of 2 there is some theoretical overestimation: the theoretical values coincide with the upper limit of the measurement values. Above a ratio of 2 the theory follows the trendline of the measurement data.

- Reproduction of $T_{20}$ (according to the theory) starts to improve at an initial peak level (PNR – INR) of 50 dB; $T_S$ improves from 20 dB.

- Measurements have shown an improvement of room acoustic reproduction by increasing the initial-peak level (direct/diffuse ratio). For more data further research is needed.

- According to the theoretical approach used it seems that in order to avoid the negative effect of listening room acoustics on played back acoustics (keeping the reverberation time constant, and changing only the initial direct level), in many cases an almost impractically high level of direct sound is required.

ACKNOWLEDGMENTS

The authors wish to thank Pawel Leskiewicz from the The Royal Conservatoire, department Art of Sound, in The Hague.

REFERENCES


