Spectral and perceptual properties of a transfer chain of two rooms

Andreas Haeussler
Acoustics Group, cluster of excellence "Hearing4all", Carl-von-Ossietzky University Oldenburg, Germany, andreas.haeussler@uni-oldenburg.de

Steven van de Par
Acoustics Group, cluster of excellence "Hearing4all", Carl-von-Ossietzky University Oldenburg, Germany, steven.van.de.par@uni-oldenburg.de

Summary
This work examines the perceptual properties of the transfer chain of a recording room and a playback room. Recordings in a room have acoustical properties that can be measured with a room impulse response (RIR). Assuming the user reproduces these recordings over loudspeakers in another playback room, the acoustical properties of that playback room will also influence the sound field. Mathematically this playback chain can be expressed as a convolution of two RIRs. In an analytic comparison, the double convolution of these two RIRs is analysed and compared to the single RIR. The double convolution leads to an extended reverberant tail and to fundamental changes in the shape of both the reverberant tail and onset of the RIR. These changes can further be analysed in the frequency domain. A statistical property of a single room above the Schroeder Frequency is a fixed standard deviation (STD) of the logarithmic magnitude spectrum of 5.57 dB. The convolution of two RIRs in the time domain leads to a multiplication in the spectral domain, while the logarithmic presentation leads to an addition of the logarithmic magnitude spectrum. Thus the resulting STD is a combination of the single STDs. As a result the playback chain involving two rooms leads to an increase in spectral fluctuation strength, measured by the STD of the logarithmic magnitude which increases by a factor of square-root of two. A listening test confirmed that this increase in STD leads to a perceptible change in coloration.

PACS no. 43.55.Br, 43.55.Hy, 43.60.Cg

1. Introduction

In daily life we often listen to recordings made in a reverberant environment (recording room), being reproduced in another reverberant environment (playback rooms). When an acoustic event should be reproduced in an authentic way, the loudspeaker system is not the only limiting factor. The room in which the listening takes place can also add acoustical information. In fact the reproduction is a mixture of the recording acoustics and the acoustics of the playback room. Alike concert recordings, also studio recordings have a distinct sound that should be preserved inside a playback room [1]. The chain of two rooms is also present in audio/video conference situation [2]. In this case often algorithms and microphones are specially designed to suppress reverberant information and only focus on the target speaker. But since the speaker might also be located outside the reverberant radius, room influence is still a factor of concern. Further examples could be a radio or an television broadcast, especially when the targeted sound source is not close to the microphone. Also for modern spatial audio rendering systems like wavefield synthesis or ambisonics etc., with the goal of an authentic reproduction of an acoustic scene like a concert hall, the question of the influence of the playback room arises [3]. This situation is also linked to electro acoustic systems, used for sound reinforcement in concert halls, where artificial reverberation is applied to the signal [4]. It can be said that the scenario of a second reverberant characteristic is quite usual these days. The situation with two rooms in the transfer chain will further be referred to as a room-in-room (RinR) condition. This paper will first present an objective evaluation of the room-acoustical consequences of RinR audio reproduction. Second, the perceptual consequences of a RinR audio reproduction will be presented.

2. Objective Evaluation

This section, subdivided in temporal and spectral characteristics, describes the physical properties of a convolution of a first room impulse response (RIR1) with a second room impulse response (RIR2).

2.1. Temporal characteristics of a RinR-IR

Mathematically the transfer chain of two rooms is a convolution [5] of the RIR1 of the recording room and the RIR2 of the playback room resulting in a "room-in-room" impulse response (RinR-IR) [1]:

$$\mathrm{RinR-IR} = \int_{-\infty}^{+\infty} \mathrm{RIR1}(\tau) \cdot \mathrm{RIR2}(t-\tau)\mathrm{d}\tau \quad (1)$$

$$= \mathrm{RIR1} * \mathrm{RIR2}. \quad (2)$$

Equation 1 implies that each pulse of the RIR1 of the playback room evokes the temporal characteristic of the RIR2 of the recording room, leading to a denser temporal structure. Derived from a mirror source model [6], the temporal structural density n of the arriving reflections can be expressed as:

$$n = \frac{\mathrm{d}N}{t} = \frac{4\pi c^3 t^2}{V}, \quad (3)$$

where the total number of reflections until timepoint t is given by N which is equal to:

$$N = \frac{4\pi c^3 t^3}{3V}. \quad (4)$$

To gain insight into the convolution of two RIRs, the resulting number of reflections N_{RinR} could be easily derived by multiplying the total number from each RIR.

$$N_{RinR} = N_1 \cdot N_2 \quad (5)$$

Thus this equation does not allow to predict the number of reflections until a specified time t. For that purpose, equation 4 has to be reconstructed. The number of reflections for a single RIR at a time t for an infinitesimal slice of the RIR is:

$$\mathrm{d}N_1(t) = n_1(t) \cdot dt \quad (6)$$

Since every impulse of RIR1 evokes the hole temporal structure of RIR2, the number of pulses that are excited until time point τ in RIR2 by a single pulse in RIR1 is given by $N_2(\tau - t)$. The total number of pulses is given by the integral:

$$N_{\mathrm{RinR}} = \int_0^{\tau} \mathrm{d}N_1(t) \cdot N_2(\tau - t) \quad (7)$$

$$= \int_0^{\tau} n_1(t) \cdot N_2(\tau - t)\mathrm{d}t \quad (8)$$

When calculating the number of pulses N for a RinR reproduction, the resulting number of pulses can be seen in the equation below:

$$N_{\mathrm{RinR}} = \frac{4}{45}\pi^2 \frac{c^6}{V_1 \cdot V_2} \cdot t^6. \quad (9)$$

Compared to equation 4 a clear increase in number of reflections can be seen, which leads to a general different temporal structure of a RinR-IR.

To visualise the effect, two RIRs are convolved with each other, namely RIR1 ($T_{30} = 550$ ms) and a RIR2 ($T_{30} = 1200$ ms), with the receiver position outside the reverberant radius. The representative RIRs are shown in figure 1. To make the effects more visible, a logarithmic presentation and a exponential smoothing with a time constant of 2.5 ms is used [7]. RIR1 and RIR2 have a step climax at the start, followed by the early reflections and the linear decay of the diffuse reflections. Compared to RIR1 and RIR2, the resulting RinR-IR shows an extended reverberant tail. Also the linear behaviour of the diffuse reflections does not hold, given that a certain convexity is visible. While comparing the onset of the RinR-IR with the underlying RIRs, it seems that the energy of the early reflections increases in case of the RinR situation and that the direct sound component is relatively weak in comparison to the early reflections. The RinR-IR has a certain transient distortion. The temporal structure of the RinR-IR seems to differ in temporal density of the pulses and in the energy distribution. Since changes in the temporal domain and the spectral domain are linked together, the spectral properties, which are associated to coloration [8], are influenced too. The next chapter focuses on the spectral shape and properties of the RinR-IR.

2.2. Spectral characteristics of a RinR-IR

Sound played inside a room introduces reverberation to the signal. The reverberant structure from the source to the microphone can be described over the RIR in the time domain, corresponding to the transfer function in the frequency domain. For an analytic presentation of the frequency content the logarithmic magnitude spectrum is used:

$$P_H = 20 \cdot log_{10}(|H|). \quad (10)$$

With the paper of Schroeder 1954 [9] a statistical room approach for the description of the spectrum has been introduced. The underlying assumption is that the sum of plane waves with random phases and amplitudes, with reflections arriving from random directions, at frequencies with a high modal overlap, is a random process. This frequency area is defined above the Schroeder Frequency [10]. For the stochastic model the real and the imaginary part of the the

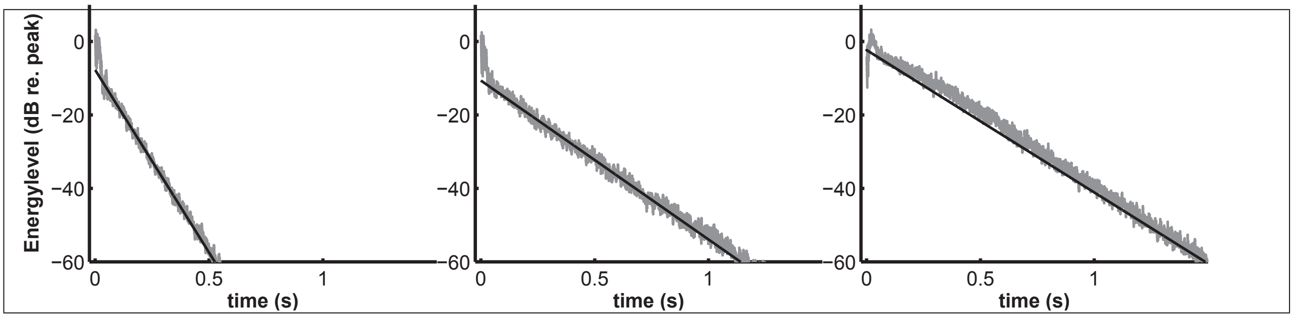

Figure 1. The logarithmic presentation of the RIR of room 1 and room 2 (left and middle) and the resulting RinR-IR (right). The black line approximates the linear decay of the diffuse reflections.

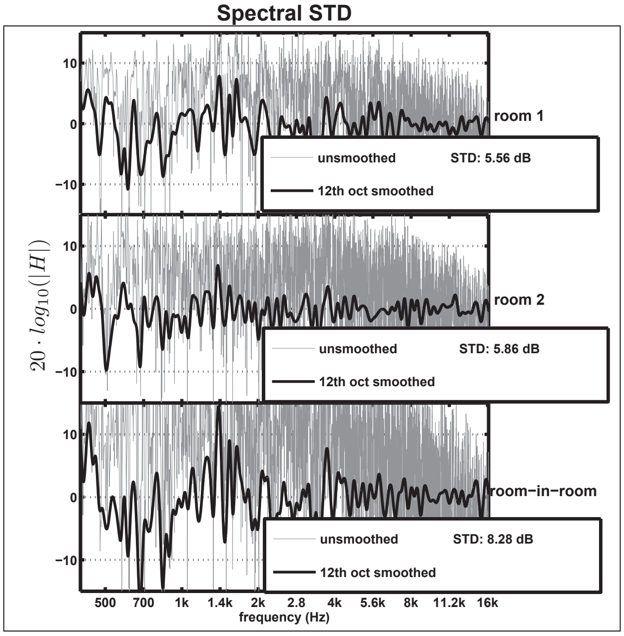

Figure 2. The logarithmic magnitude spectrum of the RIR1 (top) and RIR2 (middle) and the resulting RinR spectrum (button). Illustrated in the thin grey line is the unsmoothed spectrum, the thick line represents the $\frac{1}{12}$-oktave smoothed spectrum. Additionally the STD of the unsmoothed spectrum is given.

complex spectrum are considered as independent normal distributed processes, that can be expressed as a logarithmic magnitude response, which has a magnitude probability distribution resulting from a logarithmically transformed Rayleigh distribution [11]. In figure 2 (top and middle panel) a magnitude spectrum of a RIR can be seen. The reverberation inside a room introduces a fluctuation around a mean value that can measured over the standard deviation (STD). The result of the analysis by Schroeder is that this STD of a single room is equivalent to 5.57 dB above the Schroeder Frequency, which will here be rounded up to a value of 5.6 dB. The theoretical assumption is also only valid outside the reverberation radius, thus the STD decreases inside the critical distance [12].

As seen in the last chapter a convolution of two singe RIRs leads to a different temporal structure in the RinR-IR, in particular the increase in energy in the early reflections. Since changes in the statistical time domain structure leads to a alternation of the statistical properties in the frequency domain, the magnitude spectrum of the RinR-IR is focus of this chapter. In figure 2 the logarithmic magnitude spectrum of the same two single RIR are shown (top and middle panel) as used in figure 1. The resulting logarithmic magnitude spectrum of the RinR-IR is shown below. The unsmoothed spectrum is given by the grey line. Also a smoothed version ($\frac{1}{12}$- octave) is given by the thick line to allow easier interpretation of the spectral fluctuations. As an example the dip at 700 Hz increases as well as the peak around 1.4 kHz. The smoothed line shows that the fluctuation for the RinR situation increases compared to the single rooms.

For an analytic interpretation of the unsmoothed spectrum the STD is calculated above the Schroeder Frequency. Since the Schroeder Frequency is unknown for the RinR-IR, the lower corner frequency is chosen rather high at 400 Hz, as compared to the Schroeder Frequencies of the single rooms (121 Hz, 122 Hz). It can be seen that the single rooms have a STD of 5.56 dB (room 1) and 5.86 dB (room 2), as to be expected from the calculated STD of Schroeder. The STD of the RinR logarithmic magnitude spectrum has clearly increased with respect to the single room to a value close to 8 dB.

The theoretical impact of the convolution of two single RIR to the STD of the spectrum can be shown over the addition of the logarithmic spectrum:

$$h_{\mathrm{RinR}} = h_1 * h_2 \quad \circ\!\!-\!\!\bullet \quad H_{\mathrm{RinR}} = H_1 \cdot H_2$$
$$log(|H_{\mathrm{RinR}}|) = log(|H_1|) + log(|H_2|). \tag{11}$$

The STD of $log(|H_{RinR}|)$ can be calculated over the single STD of the logarithmic magnitude spectrum $log(|H_1|)$ and $log(|H_2|)$

$$STD^2_{\mathrm{RinR}} = \mathrm{STD}^2_1 + \mathrm{STD}^2_2 + 2 \cdot r_{1,2} \cdot \mathrm{STD}_1 \cdot \mathrm{STD}_2, \tag{12}$$

while $r_{1,2}$ represents the correlation coefficient of the logarithmic magnitude spectrum of the single

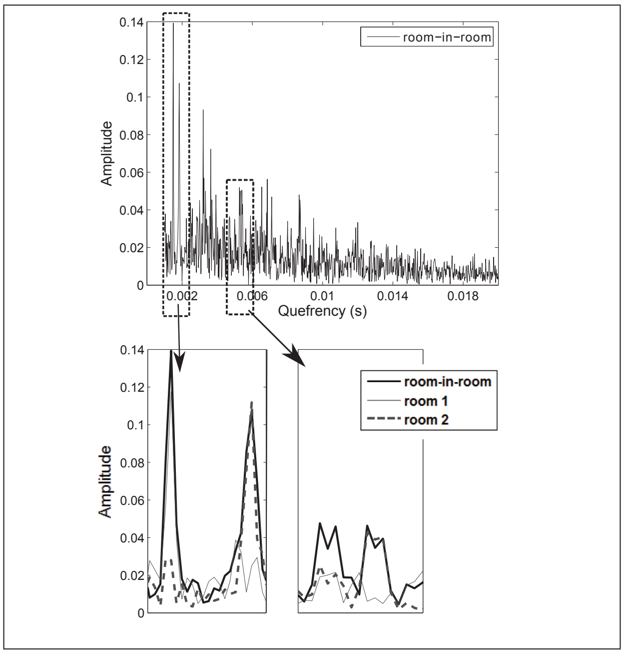

Figure 3. Cepstrum of the RinR-IR (above) and an enlargement of certain sections (below), where also the cepstrum of RIR1 and RIR2 are shown.

RIR. For a general formulation, it is assumed that the STD of a single room is around 5.6 dB and the spectra are uncorrelated. The formulation from equation 12 can be simplified to:

$$\mathrm{STD}_{\mathrm{RinR}} = \sqrt{2} \cdot \mathrm{STD}_{\mathrm{room}} = \sqrt{2} \cdot 5.6 \text{ dB} \qquad (13)$$

Thus it can be seen that in the RinR-IR the spectral fluctuation is increased by a factor of $\sqrt{2}$ as compared to RIR of a single room. We will now present a cepstral representation of the RinR-IR to analyse the rate of fluctuation. The cepstrum is defined as:

$$Ceps = \mathcal{F}^{-1}(P_H) \qquad (14)$$

and allows to analyse the fluctuation strength more closely. In figure 3 the cepstrum of the beforehand used RinR-IR from figure 1 and 2 is shown. Lower quefrencies describe the rapid frequency oscillation (quefrequency $= 0.001 \triangleq 1000$ Hz) and higher quefrencies describe the slower oscillations (quefrequency $= 0.02 \triangleq 50$ Hz). Interesting to note is mainly the composition of the peaks in the cepstrum of the RinR-IR, which is shown in the inset, together with the cepstrum of RIR1 and RIR2.

In the left lower figure can be seen, that the RinR-cepstrum is composed of the single cepstral peaks of the single RIRs, namely the first peak is a property of RIR1 and the second peak is property of RIR2. The right lower figure also shows that the RinR amplitude of the left peak is a composition of RinR1 and RinR2. Therefore the amplitude of the quefrencies of a RinR-cepstrum is a composition of the single cepstras. This is due to the linear properties of the fourier analysis of the logarthmic spectral presentation:

$$Ceps = \mathcal{F}^{-1}\Big(log(|H_1|) + log(|H_2|)\Big)$$
$$Ceps = \mathcal{F}^{-1}\Big(log(|H_1|)\Big) + \mathcal{F}^{-1}\Big(log(|H_1|)\Big). \qquad (15)$$

This analysis states, that the amplitude of the fluctuations in the spectrum gain strength and is an addition of the cepstral properties of the single cepstras. There does not seem to be a general shift in the cepstrum along the quefrency axis, suggesting that the general rate of fluctuation does not change due to the RinR condition as compared to a single RIR.

In summary, the spectrum of a RinR-IR has different statistical properties compared to a single room. The peaks in the spectra increase, as the STD can be used to quantify these changes. The under normal circumstances measured STD of 5.6 dB of the logarithmic spectrum of a single room increase about the square root of two.

3. Subjective Evaluation

As seen in the previous chapter, it is characteristic for a single room to have a STD of the logarithmic magnitude spectrum around 5.6 dB. While convolving two RIRs with each other, this STD is influenced and increases to about $\sqrt{2} \cdot 5.6$ dB. Several studies [8] [13] have used the STD of a spectrum as an aid to predict the perceived coloration. To investigate, if these objective changes in the spectral domain lead to a change in the perceived coloration, the following listening experiment has been designed.

3.1. Method

The basic idea behind the experiment is to have a RinR-condition to listen to and a single room where the coloration can be adjusted. To adjust the coloration the processing from the block diagram in figure 4 has been applied. The first 50 ms where extracted and transferred into the logarithmic spectral domain. Here a faktor k is multiplied with the logarithmic transfer function, which allows to increase the spectral content if $k > 1$, a reduction with $k < 1$. The resulting spectra is transferred back into the time domain, resulting in a zero phase filter.

This processing was also applied to the RinR-condition with $k = 1$. Important to note is that the spectral composition of the spectra remains, however the temporal structure of the RIR is no longer apparent. Listening participants were asked to match the amount of coloration of a RinR combination. The STD of the adjusted RIR is used as an estimate of the perceived coloration. In table I the used rooms are listed.

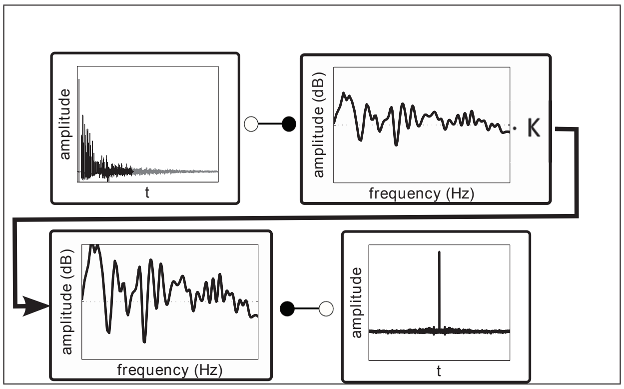

Figure 4. Block diagram to adjust the spectral content of the spectrum

Table I. Table with the physical measures and room data of the single rooms for the coloration experiment

room 1	[b = 5 m, l = 10 m, h = 3 m], V = 150 m^3 $d_1 = 5$ m, $T_{60} = [0.1\ \ 0.5\ \ 1\ \ 4]\ s$
room 2	[b = 6 m, l = 7 m, h = 3 m], V = 126 m^3 $d_2 = 3$ m, $T_{60} = [0.1\ \ 0.5\ \ 1\ \ 4]\ s$
test-room	[b = 6 m, l = 7 m, h = 3 m], V = 126 m^3 $d_3 = 4$ m, T_{30} = adjusted to the RinR-condition

The test-room varies in reverberance (change in T_{30}-time), while the position, shape and volume are kept constant, to match the influence of reverberation energy from the test-room to the RinR combination. Two different rooms where used to generate the RinR-combinations (see again table I). They basically differ in volume and the distance from source to microphone, Room 1 with a distance of 3 meters and room 2 with a distance of 5 meters. At this position the T_{30}-time was adjusted to the four conditions, resulting at a T_{30}- time of 100 ms, 500 ms, 1 s and as an extreme case 4 s at the microphone. This single rooms were convolved with each other, altogether resulting in 16 combinations. All rooms were generated in a simulated environment [14] and were presented under diotic conditions. Another simplification has been used while leaving the absorption coefficients constant over the frequencies. These rooms can again be compared with the size of a normal living room/ recording room/ lecture room. In this experiment a dry piano recording has been used as musical instrument (4 s), while the played contend was ongoing, with no breaks in between, but also a little transient structure due to the change in chords. When comparing the zero phase filter convolved with the piano and the real RIR convolved with the piano the sound differed only minimal, cause the ongoing sound of the piano masked most of the temporal structure. The experiment was conducted in a single wanded listening

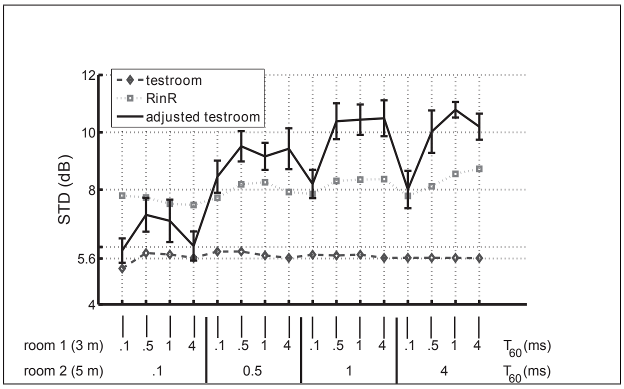

Figure 5. The result of the coloration matching experiment. To quantify the results, the STD of the logarithmic magnitude spectrum has been used

cabin over headphones (Sennheiser HD650), with an listening rms level of 65 dB SPL. In this experiment, ten listening subject participated, all of them with an acoustical background. Only one run was conducted, without any repetition and retest.

3.2. Subjective results

For an easier interpretation of the listening results, the STD of the logarithmic magnitude spectrum is used . In this case the broadband STD of the spectrum was analysed starting from 80 Hz up to 13000 Hz, disregarding the Schroeder Frequencies. The results can be seen in figure 5, the resulting STDs in dB of the test-room, the RinR and the adjusted test-room. On the x-axis, the two room conditions are listed. Room 1, with a 3 meter distance from source to receiver and room 2 with a distance of 5 meters are combined to the RinR-condition and the single T_{30}-times of the rooms are denoted. Listening participants adjusted the coloration strength of a test-room to the perceived coloration strength of a RinR-condition. The resulting spectra where also analysed with the STD and are denoted with the black ongoing line. For the adjusted test-room the standard error is derived from the mean STD over all subjects. The first quarter of results, the combination of all reverberation times of room 1 with room 2 at rather theoretical T_{30}-time of 100 ms, has a slight increase in STD. While increasing the T_{30}-time of room 2 to 0.5 s, the resulting STD of the adjusted room increased. For the third quarter, the STD of the adjusted test-room, increased even further, especially for the condition with a T_{30}-time of room 1 above 100 ms.

3.3. Discussion

To compare the coloration strength between a single room and a RinR-condition a new listening test design has been developed. Primarily it takes the early reflections into account and neglects the temporal reverberant structure of the RIR. When examining the STD of the test-rooms it can be seen that despite the different reverberation times, the STD stays relatively constant with a mean STD of 5.7 dB over all conditions. The RinR-condition increases in STD as to be expected with a mean STD of 8.0 dB, which verifies the increase with a "square root of two" thumb rule. The results of the listening test show a general increase in the perceived coloration strength for RinR conditions as compared to single rooms. The influence of the low reverberant room 2 with a 5 meter distance in the first quarter, shows less drastic increase in coloration, rather a minor change. When room 2 has a T_{30}-time of 500 ms or higher, the increase in coloration is apparent. It is interesting to note, that some combinations of the T_{30}-time are available twice. For example room 1 ($T_{30} = 1$ s) and room 2 ($T_{30} = 0.1$ s) can be compared to room 1 ($T_{30} = 0.1$ s) and room 2 ($T_{30} = 1$ s). Still the same T_{30}-time is apparent at the microphone, the adjusted STD still is clearly different. There is no good explanation for this. Note also that for the RinR condition the adjusted STD of the matching room differs from the STD of the RinR-IR. Thus it seems that the STD of a room as such is not a good predictor of the perceived coloration strength. Overall it can be said that a combination of rooms, with a reverberation around 500 ms or higher increases the coloration strength.

4. CONCLUSIONS

In this study a convolution of two rooms has been discussed. It has been shown that the temporal characteristics of a RinR-condition differs from the underlying RIR. The number of pulses increases, while the energy distribution changes. The onset/transient of the RinR-IR has certain smoothing of the transient. The energy rises and seems to cumulate in the region of the early reflections, also the linear decay of the later reflections does not hold. In the frequency domain, the STD has been used to quantify changes due to the convolution. Since a normal RIR has a STD of around 5.6 dB in the spectrum, the RinR-conditions leads to an increase of $\sqrt{2}$. A new method to quantify coloration has been introduced, with which it could be shown that for most conditions an increase in perceived coloration has been found.

Acknowledgement

This project was funded by the PhD program "Signals and Cognition" (Niedersächsisches Ministerium für Wissenschaft und Kultur). The authors like to thank Torben Wendt [14] for the help of deriving equation 9.

References

[1] Hak C.C.J.M.,Wenmaekers R.H.C.: *The Impact of Sound Control Room Acoustics on the Perceived Acoustics of a Diffuse Field Recording.* WSEAS Transactions on Signal Processing 6.4 (2010): 175-185.

[2] Svensson, U. Peter, HE-B. Zidan, and Johan L. Nielsen: *Properties of convolved room impulse responses* Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on. IEEE, 2011.

[3] Grosse, J.; van de Par, S., *Perceptually accurate reproduction of recorded sound fields in a reverberant room using spatially distributed loudspeakers*, Selected Topics in Signal Processing, IEEE Journal of , vol.PP, no.99, pp.1,1

[4] Svensson, U. Peter: *Energy-time relations in a room with an electroacoustic system.* The Journal of the Acoustical Society of America 104.3 (1998): 1483-1490.

[5] Vogel, Peter, and Diemer de Vries.: *Electroacoustic system response in a hall: a convolution of impulse sequences.* Journal of the Audio Engineering Society 42.9 (1994): 684-690.

[6] Kuttruff H. (2009) *Room Acoustics* Spon Press - Fifth Edition

[7] S. Weinzierl (2008) *Handbuch der Audiotechnik* Springer-Verlag Berlin Heidelberg

[8] Meynial, X., and O. Vuichard. *Objective measure of sound colouration in rooms.* Acta Acustica united with Acustica 85.1 (1999): 101-107.

[9] Schroeder M. : *Die statistischen Parameter der Frequenzkurven von grossen Räumen* Acustica 4 (1954) 594-600

[10] Schroeder, M. R., and K. H. Kuttruff. *On frequency response curves in rooms. Comparison of experimental, theoretical, and Monte Carlo results for the average frequency spacing between maxima.* The Journal of the Acoustical Society of America 34.1 (1962): 76-80.

[11] H. Cramer *Mathematical Methods of Statistics* Princeton University Press, Princeton, NJ, 1946,chap. 28.6

[12] Jetzt, J. J. *Critical distance measurement of rooms from the sound energy spectral response.* The Journal of the Acoustical Society of America 65.5 (1979): 1204-1211.

[13] Georganti, E., Dillier, N, and Mourjopoulos, J. *Measuring perception of coloration due to early reflections in binaural room impulse responses.* 7th Forum Acousticum. 2014.

[14] Wendt, Torben; van de Par, Steven; Ewert, Stephan D. (2014): *A Computationally-Efficient and Perceptually-Plausible Algorithm for Binaural Room Impulse Response Simulation*, JAES Volume 62 Issue 11 pp. 748-766; November 2014