Synthesis of room impulse responses based on simulated energy decay curves

Lukas Aspöck[1], Michael Vorländer[1]

[1] *Institute of Technical Acoustics, RWTH Aachen University, 52074 Aachen, E-Mail:las@akustik.rwth-aachen.de*

Introduction

Traditionally, room acoustic simulation software is used to calculate room acoustic parameters of a room. For this purpose, it is sufficient to estimate the energy decay using a reduced temporal resolution, which at least allows the calculation of energy decay rates (e.g., for the calculation of the EDT or T30) and energy ratios (such as the clarity or definition). For auralizations however, it is required to synthesize a room impulse response (RIR) corresponding to the room acoustics simulation results. This RIR should have an adequate temporal resolution, e.g., a sampling rate of 44.1 kHz. In geometrical acoustics, the early part of the RIR is typically calculated by an image source model, which directly provides the required temporal information. The late part is usually based on a predicted energy decay curve and does not include exact information (time and pressure) about the incoming reflections.

This work investigates configurations for the synthesis of the later part of a RIR, including the variation of the reflection model, e.g. its statistical distribution, and the temporal resolution of energy decay curve. The resulting impulse responses are generated with the goal to maintain the values of simulated room acoustic parameters and at the same time provide RIRs which are perceptually indistinguishable from corresponding measured RIRs if applied in auralization experiments. This is the groundwork for a detailed validation of simulation results and a perceptual evaluation of simulated RIRs answering questions regarding the authenticity and plausibility of room acoustics simulations.

Synthesis of room impulse responses

For the simulation of the reverberation tail using geometrical acoustics methods, ray or beam tracing methods usually provide the energy decays by generating energy histograms (see left part of Figure 1). If room acoustics are predicted just by application of the Sabine or Eyring reverberation time equation, the energy decay is assumed to be perfectly exponential, according to the statistical reverberation theory. Based on this description of the room's energy decay, a broadband RIR has to be synthesized if a more detailed parameters analysis or an auralization is planned. This work focusses on the synthesis of the later part of the RIR (see right part of Figure 1, colored in red).

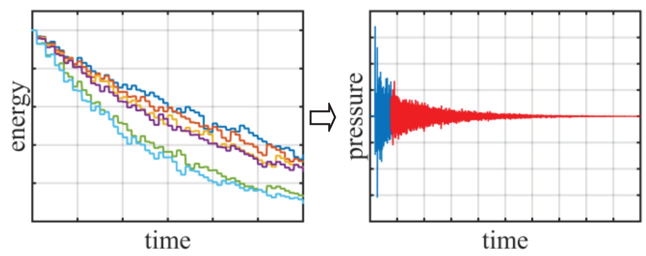

Figure 1: From energy histograms for separate frequency bands to broadband room impulse response.

In 1979, Moorer presented a reverberation synthesis based on frequency dependent reverberation times [1]. Stationary white noise was generated and processed by an octave band filter bank. The band-passed noise was then multiplied with an exponential decay function according to the reverberation time of the corresponding frequency band. The resulting exponentially decaying impulse responses were then summed up and in combination with a direct sound impulse, used for auralization. Although white noise fulfills the requirement of a flat frequency spectrum, it does not account for the increasing reflection density of a real RIR and thus would only be valid after a certain time limit [2]. To describe how the reflection density increases over time in a room with a volume *V*, Cremer derived an equation (see Eq. 1) based on the image sources of a rectangular room [3].

$$N_r(t) = 4\pi \frac{c^3 t^2}{V} \qquad \left[\frac{1}{s}\right] \qquad (1)$$

Kuttruff showed that the equation also applies to rooms with arbitrary shape [4]. The equation can be applied in the synthesis process of the RIR: For each sample, a uniformly distributed random process is used to decide whether a reflection occurs during that time or not. Another proposed modification of Moorer's model is the usage of reflection sequences individually for different frequency bands. This allows the modified application of Eq. 1 with regards to frequency dependent effects such as surface scattering.

Evaluation of the reflection density in rooms

Due to limited processing capacity, the reflection density modelling often included an upper limit ranging from $250\frac{1}{s}$ up to $10{,}000\frac{1}{s}$ [5]. The higher values can be found in more recent publications such as [6]. Synthesized RIRs using this limit were described as being free from flutter or rattling [7]. Figure 2 shows the reflection density function according to Eq. 1 for three rooms with a different room volume.

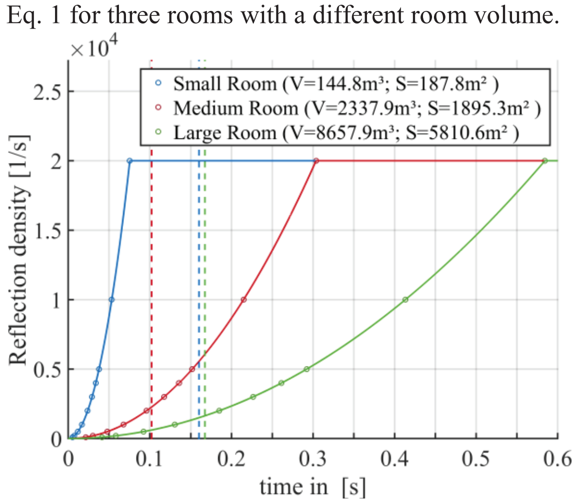

Figure 2: Reflection density function for different room sizes, with a maximum reflection density of 20,000 1/s. Dashed lines show the Hidaka mixing time for each room.

The graphs were generated for an upper limit of 20,000 reflections; circles indicate additionally investigated maximum densities starting from 100 reflections per second. The comparison of the rooms shows that the reflection density quickly increases for the small room, reaching the maximum of 20,000 within the first 100 ms of the RIR, whereas this density is reached at almost 600 ms in case of the large room. It should be noted that this graph does not describe the total amount of reflections up to this point, but the amount of reflections which would be occurring over the period of 1 s.

The dashed lines indicate the mixing time, based on a predictor proposed by Hidaka [8]:

$$t_{Hidaka} = 0.08 * RT_{500Hz} \quad [s] \qquad (2)$$

The mixing time is usually considered as the point in time of a RIR, when the transition from early reflections to the reverberation tail takes place. For this work, Hidaka's equation was chosen as it accounts for the reverberation time and is not just based on geometrical properties of the room. Additionally, in comparison to other equations (e.g., by Rubak or Cremer), Hidaka's equation leads to the highest mixing times for all three rooms and thus represents a more conservative estimation of the mixing time.

The mixing time is relevant for the investigation of the reflection density as the pressure distribution of the reverberation tail follows a Gaussian distribution once the mixing time is passed [9]. Thus, checking if the reverberation tail takes on a Gaussian distribution after this point in time represents an objective method to check the chosen reflection density function and/or its upper limit.

The mixing times of the three investigated rooms show that the small room is more reverberant than the medium room due to its substantially higher mixing time. Although the small and the large room differ in volume, their reverberation time in the 500 Hz octave band is similar, also leading to almost identical mixing time. In contrast to the medium and large room, for the smaller room the maximum reflection density of 20,000 reflections is reached before the mixing time. Thus, for the two larger rooms, the sound field can be regarded as being fully mixed, although the current reflection density has not reached its upper limit, but values around 2,000 reflections per second.

Evaluation methods of the reverberation tail

Because the perceptual impression, when listening to auralizations based on RIRs, is mostly dominated by direct sound and early reflection, the evaluation and analysis of the reverberation tail is often only conducted on a very basic level. Here, the first step would be to check if the energy decay curve (EDC) matches the desired curve. The next step of the analysis is typically the evaluation of room acoustic parameters such as the clarity or the reverberation time. While avoiding deviations when comparing EDCs and room acoustic parameters to measured results is essential for a correct synthesis of RIRs, eventually the goal of the synthesis procedure should result in simulated RIRs which are indistinguishable in listening experiments when compared to thoroughly measured results.

The synthesis however, has a lot of configuration possibilities:

- Frequency band resolution (e.g., in third octaves)
- Type and order of bandpass filter for filter bank
- Temporal distribution of impulse sequences (uniform or Poisson distributed [6][7])
- Temporal resolution of simulated or measured energy histogram
- For the synthesis of binaural RIRs: Integration of spatial information into the reverberation tail [10]
- Reflection density model and reflection density limit

Because conducting listening experiments for all these synthesis configurations is an extensive research task, a more detailed objective analysis of the synthesized RIRs in addition to decay rates and room acoustic parameters is desired.

One option is the statistical evaluation of the temporal structure of the RIR. To check if the sound pressure follows a Gaussian distribution, one could conduct empirical distribution tests such as the *Lilliefors* test. Other options are moment tests, e.g., a kurtosis evaluation. The normalized kurtosis of a sound pressure vector x is calculated according to Eq. 3:

$$K_{norm} = \frac{E(x-\mu_x)^4}{\sigma_x^4} - 3 \qquad (3)$$

In room acoustics, the (normalized) kurtosis has been used to determine the mixing time in RIRs [11] or to evaluate the diffuseness of measured RIRs [12]. The kurtosis is a measure to describe the amount of extreme outliers, in case of an RIR the amount of reflections with an irregular sound pressure:

- $K_{norm} \gg 0$ High amount of irregular reflections
- $K_{norm} \approx 0$ Gaussian distribution
- $K_{norm} < 0$ Uniform distribution

Evaluation of synthesis result for three rooms

According to the described synthesis procedure, RIRs for three rooms (see Figure 2) were generated and compared to the corresponding measured RIRs of these rooms. To eliminate potential simulation errors, the synthesized results were not based on simulation results, but on the measured energy histograms, evaluated for timeslots with a length of 5 ms for octave bands between 63 Hz and 16 kHz.

A large set of synthesized RIRs were generated based on varying parameters (see above). In this paper, the evaluation for different maximum reflection densities is presented (from 100 up to 10,000 reflections per second). The evaluation only relates to the late part of the RIR: A time window starting at 150 ms was applied to measured and synthesized RIRs.

The results of the traditional evaluation based on EDCs and T20 showed no substantial differences of measured and synthesized RIRs. These results were expected as the synthesis was based on the original energy histograms. Only for a third octave band evaluation of T20, deviations above the JND of 5%, especially in the lower frequency bands, can be observed. These effects however do not correlate with the maximum reflection density chosen for the synthesis process, but with the random distribution of the included reflections.

A kurtosis evaluation was conducted for the same measured and synthesized RIRs. According to similar research [5][9], the normalized kurtosis was evaluated using a sliding window with a length of 1024 samples.

Figure 3 shows an example for the kurtosis evaluation for the large room. Three different synthesized RIRs are compared to the normalized kurtosis of the measured RIR between 150 and 250 ms.

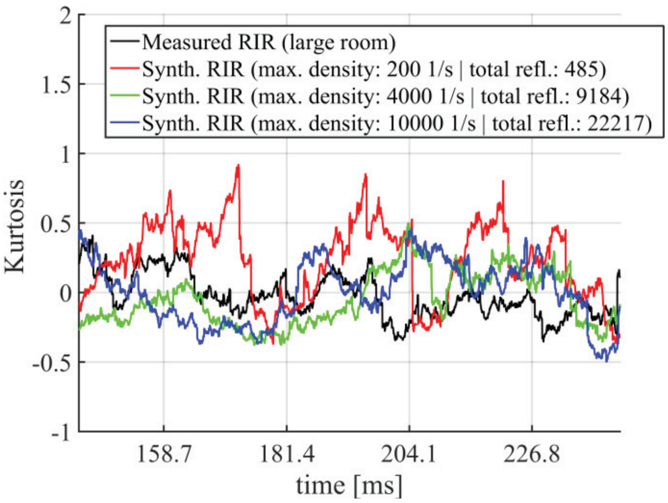

Figure 3: Normalized kurtosis evaluation for four RIRs of the large room. Evaluation is done with a sliding window (1024 samples) starting at 150 ms (Hidaka mixing time: 167 ms). Plot shows the curve for a measured RIR and three synthesized RIRs for three different maximum reflection density limits (200, 4,000, 10,000)

It can be observed, that the normalized kurtosis of the measured RIR (black curve) is close to zero and thus mostly follows a Gaussian distribution, while the synthesized RIR with a substantially reduced amount of reflections cannot be described as Gaussian distributed for most time windows (red curve, maximum reflection density 200 $\frac{1}{s}$, in total 485 reflections for the full length RIR of 2.5 s). For a higher reflection density (green curve, 4,000 $\frac{1}{s}$) the kurtosis values are higher in comparison to the measurement, but vary around the same amount. The further increment of the maximum reflection density does not show a substantial improvement.

The total evaluation 10 RIRs for all three rooms shows that the average difference between the kurtosis of measured and synthesized RIRs converges towards zero for maximum reflection densities of $1{,}000 - 2{,}000\ \frac{1}{s}$ or higher. Thus in these cases, the synthesized RIRs take on a Gaussian distribution. This finding confirms suggestions for sufficient reflection densities in synthesized RIRs. This evaluation however only covers aspects regarding the statistical pressure distribution of the reverberation tail. First listening experiments indicate that indeed maximum reflection densities of up to 10,000 reflections per second are required for the synthesis of RIRs which are perceptually indistinguishable from measured RIRs.

Summary and Outlook

A model for the synthesis of room impulse responses based on energy decay curves has been presented. An evaluation and comparison to measurements was conducted for three different rooms with volumes ranging from 200 m³ to 8000 m³. This paper discussed the applied reflection density functions and investigated the impact of the synthesis based on different maximum reflection density values. The traditional EDC and T20 evaluation showed rather small influence of the maximum reflection density on the results while a kurtosis evaluation showed that at least a reflection density of around 2,000 reflections per second is required to achieve a Gaussian distribution of the same quality as the measured RIRs. If however, the synthesized RIRs are compared to measured RIRs, first informal listening experiments suggest using higher reflection densities, despite deviations of evaluated room acoustic parameters being minimal for lower maximum reflection densities.

The presented results only cover a small aspect of the research related to the synthesis of room impulse responses. An analysis of different temporal resolution of energy histograms (input for the synthesis) has been conducted and concepts for adjusting the reflection density function for rooms with inhomogeneous absorption distribution and accounting for surface scattering have been prepared. Further investigations include an extended analysis of room acoustic parameters as well as a spectral comparison of measured and synthesized RIRs.

Due to limitations of measurement techniques, it is a challenge to count reflections in measured RIRs. To create better reference data, a RIR based on an image source simulation for a high order will be generated. Additionally results of a ray tracing algorithm will be analysed to validate the applied reflection density function for regular and irregular room shapes.

Acknowledgments

This research was funded by the DFG (German Research Foundation) as part of the SEACEN research unit (www.seacen.tu-berlin.de).

References

[1] Moorer, J. A. About this reverberation business. Computer music journal, 1979, pp. 13-28

[2] Jot, J. M., Cerveau, L., & Warusfel, O. Analysis and synthesis of room reverberation based on a statistical time-frequency model. In Audio Engineering Society Convention 103, 1997

[3] Cremer, L., & Müller, H. A. Die wissenschaftlichen Grundlagen der Raumakustik. Hirzel Verlag, Stuttgart, Germany, 1978

[4] Kuttruff, H. Room acoustics. Spon Press, New York, USA, pp. 106-113, 2009

[5] Lindau, A., Kosanke, L., & Weinzierl, S. Perceptual evaluation of model-and signal-based predictors of the mixing time in binaural room impulse responses. Journal of the Audio Engineering Society, 60(11), 2012, pp. 887-898

[6] Schröder, D. Physically based real-time auralization of interactive virtual environments. Logos Verlag, Berlin, 2011

[7] Heinz, R. Binaural room simulation based on an image source model with addition of statistical methods to include the diffuse sound scattering of walls and to predict the reverberant tail. Applied Acoustics, 38(2-4), 1993, pp. 145-159

[8] Hidaka, T., Yamada, Y., & Nakagawa, T. A new definition of boundary point between early reflections and late reverberation in room impulse responses. The Journal of the Acoustical Society of America, 122(1), 2007, pp. 326-332.

[9] Abel, J. S., & Huang, P. A simple, robust measure of reverberation echo density. In Audio Engineering Society Convention 121, 2006

[10] Aspöck, L., Pelzer, S., & Vorländer, M. Using spatial information for the synthesis of the diffuse part of a binaural room impulse response. In Proc. of DAGA 2014, vol. 40, pp. 71-72

[11] Stewart, R., & Sandler, M. Statistical measures of early reflections of room impulse responses. In Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, 2007, pp. 59-62

[12] Jeong, C. H. Kurtosis of room impulse responses as a diffuseness measure for reverberation chambers. The Journal of the Acoustical Society of America, 139(5), 2016, pp. 2833-2841