A Perception-Based Parametric Model for Synthetic Late Binaural Reverberation

Philipp Stade[1,2], Johannes M. Arend[1]
[1] TH Köln, Institute of Communications Engineering, Cologne, Germany
[2] TU Berlin, Audio Communication Group, Berlin, Germany
E-Mail: philipp.stade@th-koeln.de

Introduction

The use of binaural room impulse responses (BRIRs) for the auralization of rooms is an established approach in virtual acoustics. Parametric coding of 3D-Audio is also well known in literature [1][2] and part of different standards, e.g. MPEG-4. Since large BRIR datasets are required for dynamic binaural synthesis, especially for rooms with long reverberation times, methods to reduce the amount of data are relevant for different applications like gaming or virtual reality scenarios. Furthermore it is conceivable that in some cases a plausible auralization is sufficient and high-resolution BRIRs are not needed, e.g. when using numerous BRIRs for secondary sound sources in soundscapes. In these applications, BRIRs with reduced information possibly satisfy the same as the original datasets. For this reason a scaleable system for the generation of synthetic BRIRs could be useful.
The presented investigation aims for a parametric description of the diffuse part of BRIRs. A perception-based approach is used to generate synthetic late reverberation tails. The underlying parametric model will be explained and results of the synthesis will be compared to the corresponding measured impulse responses. Furthermore a listening experiment using dynamic binaural synthesis is presented which evaluates the auralization of different rooms based on BRIRs with synthetic late reverberation tails.

System Overview

The basic idea of the approach is to reduce the late part of BRIRs (or of entire circular BRIR datasets) to their main features. It is possible to specify these features approximately with mathematical functions or discrete values - further called parameters. Perceptual thresholds should be used to omit imperceptible and hence irrelevant information of the impulse responses. This procedure leads to a substantial reduction of the BRIRs since not every sample of the late reverberation but rather only the parameters are broadcasted or stored. Moreover the model allows the adjustment, examination and variation of the parameters independently from each other.
The algorithm for the synthesis of binaural late reverberation is subdivided into two main parts. In a first step the *analysis-part* (see Figure 1) extracts the parameters from a given reference BRIR dataset. After that the *synthesis-part* (see Figure 2) uses only these parameters for the generation of the late reverberation without any knowledge about the original signal. In the proposed model, a BRIR tail is represented by the following three frequency-dependent features only:

a) energy decay curve (EDC)

b) mean energy of late reverberation

c) interaural coherence (IC).

Two inputs are needed to extract the parameters: The original measured BRIR and the desired starting sample of the synthetic reverberation tail. This time could be chosen individually, but one main idea of the model is to utilize perceptual effects of the mixing time. Therefore this moment is equally named in the diagrams and should be chosen like proposed in literature [3][4]. The perceptual mixing time denotes the moment when a listener is not able to distinguish the diffuse tail of different orientations. Therefore the diffuse tail is synthesized after the mixing time for one listener orientation only.

Analysis-Part

The analysis part is designed as follows and is shown in the block diagram in Figure 1 by means of a single BRIR. It is also possible to utilize a circular BRIR dataset for dynamic auralization. For this case the proposed model analyzes the whole dataset and computes the mean parameters over all angles.

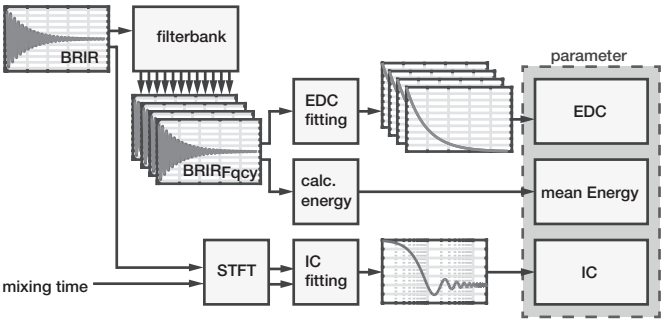

Figure 1: Block diagram of the analysis part: The parameters *EDC*, *mean Energy* and *IC* are extracted from a reference BRIR.

The analysis-part has two main stages: First the decay curve and the mean energy are calculated frequency-dependent. Therefore the reference BRIR is processed with a polyphase filter bank with near perfect reconstruction and separated into 32 frequency bands (resolution: $f < 8\,\text{kHz}$: $\frac{1}{3}$ octave, $f > 8\,\text{kHz}$: $\frac{1}{6}$ octave). The energy decay curve is approximated in each frequency band with a curve fitting algorithm to find a 6-degree polynomial that is a best fit for the logarithmic impulse response. Indeed this approach causes inaccuracies for the early part of an

Parameter	Values
Polynomial Energy Decay Curve	$(n_1+1)\cdot n_\mathrm{f}+2\cdot n_\mathrm{f}$
BRIR Length	1
Mean Energy	$2\cdot n_\mathrm{f}$
Noise Level	n_f
Polynomial Interaural Coherence	$(n_2+1)+2$
IC Length	1

Table 1: Required values for parametric model, where n_1 is the polynomial degree of the decay curve fitting, n_2 the polynomial degree of the interaural coherence fitting and n_f the number of frequency bands of the used filterbank.

impulse response, because it is not possible to represent reflections with a low-order polynomial function. But since only the late and diffuse part of the decay curve is used for the synthesis, this constraint is uncritical. To improve numerical properties of the curve fitting algorithm, a centering and scaling transformation is used. Furthermore the mean energy for the very late part of the impulse responses is calculated. The noise level in every frequency band is determined with an iterative algorithm [5] to adjust the range for the energy calculation to a desired threshold above the noise level.
In the second stage the interaural coherence (IC) of the late part of the reference BRIR is calculated. The computation of the IC is based on a short-time Fourier transform (STFT) with K frames [6]:

$$\Phi(i)=\frac{|\sum_{k=1}^{K} S_\mathrm{L}(i,k)\overline{S_\mathrm{R}(i,k)}|}{\sqrt{\sum_{k=1}^{K}|S_\mathrm{L}(i,k)|^2\sum_{k=1}^{K}|S_\mathrm{R}(i,k)|^2}}, \quad (1)$$

where S_L and S_R are the Fourier transform of the left and right channels of the BRIR, i and k denotes the indices of frequency and time. Like the decay curves, the interaural coherence versus frequency is approximated with polynomial curve fitting. Because of the more complex function of the interaural coherence, a higher-order polynomial (degree: 18) is necessary to reproduce the curve progression accurately. The used polynomial degrees for both curve fitting steps were chosen due to preliminary investigations and lead to accurate approximations of the curves. In contrast to the EDC calculation, the IC measurement is depending on the used mixing time and has to be recalculated with every new transition time. Table 1 shows the required values for the proposed model which vary with the used frequency resolution and the polynomial degree of the two fitting algorithms. Due to the used settings, 407 discrete values are needed for the synthesis of the binaural reverberation tail in this study.

Synthesis-Part

The generation of the synthetic late reverberation part based on parameters is schematically shown in Figure 2. A dual-channel white Gaussian noise with the length of the reference BRIR is generated and processed with the same filterbank used in the analysis-part. The frequency-dependent decay curves are determined with polynomial evaluation of the EDC parameters. Subsequently the white noise is shaped in every frequency band to the desired reverberation time of the reference BRIR by multiplying element-wise with the corresponding decay curves [7]. Within this procedure the spectrum of the noise is also matched to the spectrum of the original impulse responses. The mean energy of the decaying noise signals is measured frequency-dependent in the same window like the original BRIR in the analysis-stage and adjusted to the reference level. Again polynomial evaluation is used to restore the original interaural coherence function. After adding up all frequency bands, the uncorrelated decaying noise signals $d_L(n)$ and $d_R(n)$ are linearly convolved with the filters $u(n)$ and $v(n)$ to match the interaural coherence of the reference:

$$\begin{aligned} b_L(n) &= (u * d_L + v * d_R)(n) \\ b_R(n) &= (u * d_L - v * d_R)(n) \end{aligned} \quad (2)$$

where b_L and b_R are the coherence matched decaying noise signals. The filters are defined in the frequency domain as [8]:

$$U(w)=\sqrt{\frac{1+\Phi(w)}{2}},\ V(w)=\sqrt{\frac{1-\Phi(w)}{2}}. \quad (3)$$

In the last stage of the synthesis-part, original and synthesis are mixed by replacing the late reverberation part of the original BRIR with the decaying noise signals after the desired mixing time.

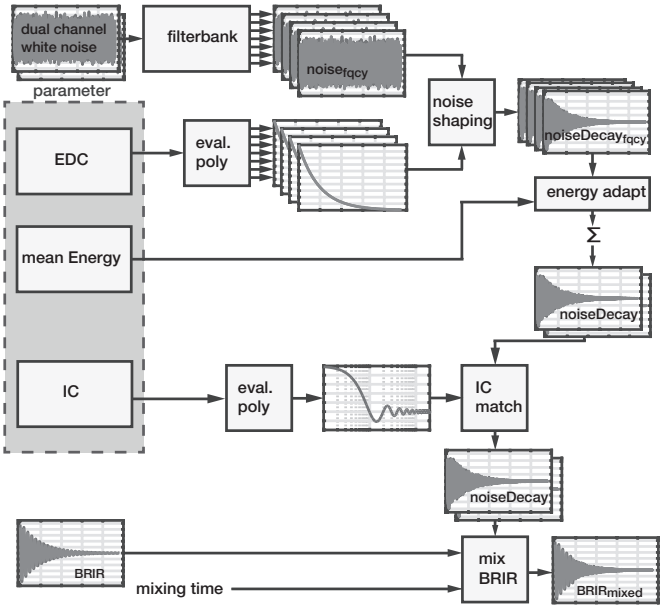

Figure 2: Block diagram of the synthesis part: The parameters *EDC*, *mean Energy* and *IC* are used to shape a dual channel white noise signal to the reference-features.

BRIR datasets synthesis

To examine the performance of the proposed algorithm, several circular BRIR datasets with synthetic late reverberation were generated and compared to the original measured impulse responses. These datasets are also used for the dynamic auralization in the listening test presented afterwards. The BRIRs of two radio broadcast studios located at the WDR Cologne (*kleiner Sendesaal*,

Rooms	Volume	Area	$RT60_{mean}$	α_{mean}
SBS	1247 m^3	204 m^2	0.83 s	0.32
LBS	6098 m^3	480 m^2	1.46 s	0.24

Table 2: Main properties of the auralized rooms [9].

named "SBS" and *KVB-Saal*, named "LBS", see Table 2) with two different sound sources each (omnidirectional, "SBC" and directional, "PAC") were synthesized at two mixing times (SBS: 20/160 ms, LBS: 40/320 ms), resulting in 8 test conditions. The mixing times were chosen according to a previous study [4]. As proposed before, the approach uses a white Gaussian noise for the synthesis of the late reverberation tail. To avoid the influence of this factor, the same noise sample was used for every synthesis. The difference between the original measured BRIR and a BRIR with synthetic reverberation based on the proposed algorithm is shown exemplary for one room (LBS SBC) in Figure 3. At both mixing times the analyzed acoustic properties are matched satisfactorily, but the differences are getting smaller at higher mixing times. Especially the early decay time and the interaural coherence depict larger differences between orginal and synthesis for the early transition time. The data volume of a circular BRIR dataset decreases by a factor of around 68 or 8 for the early or late mixing time by using the proposed approach for the synthesis of the late reverberation part.

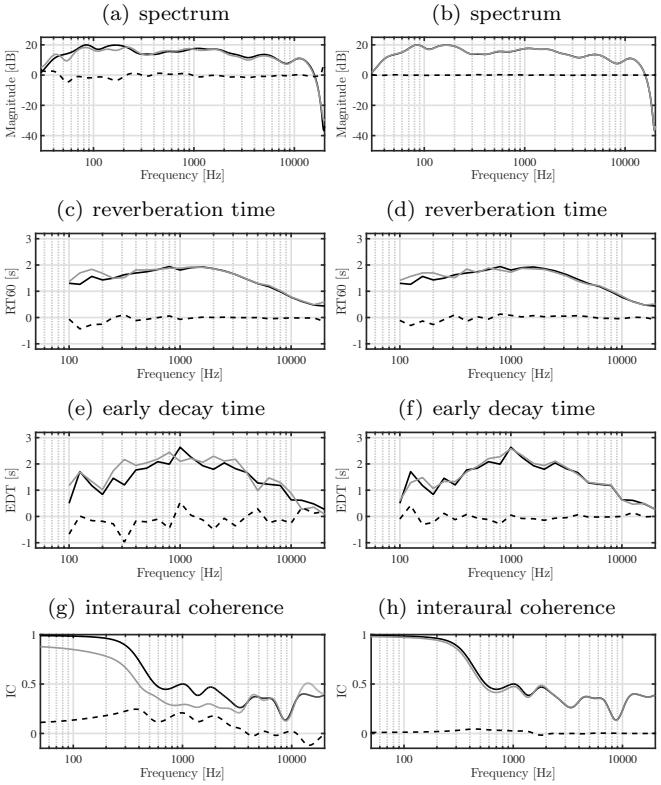

Figure 3: Differences (dotted) between original BRIR (black) and BRIR with synthetic reverberation (gray): mixing time = 40 ms (left row), 320 ms (right row)

Listening Experiments

The performance of the parametric model was evaluated in a double-blind triple-stimulus with hidden reference test paradigm according to ITU-R BS.1116-3 [10]. BRIRs with synthetic late binaural reverberation were auralized in a headphone-based virtual acoustic environment (VAE) using the *SoundScape Renderer*, *AKG K601* headphones and a *Polhemus FastTrack* headtracking system. The dynamic binaural synthesis involved the full circle horizontal plane in 1°-resolution. A series of two white noise bursts with 500 ms duration, 1 ms fade in/out (cosine squared ramp) and 1000 ms silence after the noise bursts was used as test signal and convolved with the BRIR datasets.

Twenty-two subjects aged between 20 and 45 (avg. 26) took part in this investigation. The majority (16) of them had experience in listening experiments in virtual acoustic with the used dynamic binaural system. The subjects had to identify the BRIRs with synthetic reverberation and to rate the impairment of the synthesis to the given reference (the original measured BRIR). Using a graphical interface on a tablet computer, the subjects were able play three stimuli ("A", "B" and "C") and to rate the difference of "B" and "C" to the given reference "A" with two appropriated faders. As proposed in [11] the faders were labeled only with the two attributes "identical" and "very different" (in German), but the ratings were captured continuously between 1 and 5. To improve statistical power, every condition was evaluated 10 times, resulting in 80 ratings per subject and test. The experiment included a training phase prior to the grading phase in order to make the subjects familiar with test procedure and stimuli.

The results of the listening test are shown as difference grades by subtracting the reference's rating from the the rating of the test stimulus. Negative difference grades indicate correct identifications of the test stimuli (from 0 = "identical" to -4 = "very different"). Figure 4 shows the mean difference grades with 95 % confidence intervals averaged over all 22 subjects for the four room conditions and two mixing times. T-tests (significance level α = 5 %) were used for further data analysis based on the mean values. The synthesis was always significantly detected (t-test against zero), but generally all ratings are in the upper third of the grading scale. Unsurprisingly the late mixing time was always rated significantly better than the early mixing time (two-sample t-test). Furthermore the small room (SBS) was rated slightly better than the large room (LBS), although we compensated the bigger room size with larger mixing times. In both rooms the omnidirectional source (SBC) was rated significantly worse than the directional sound source (PAC). It can be assumed that the direct-to-reverberant ratio is the critical factor: In rooms with less differences between direct sound and reverberation, the synthesis is identified clearer. To give a better overview, Figure 5 shows the difference grades of all subjects and trials for the 8 conditions in form of a boxplot. The whiskers are defined as Q3+1.5∗IQR or Q1−1.5∗IQR (Q3: upper quartile, Q1: lower quartile, IQR: interquartile range).

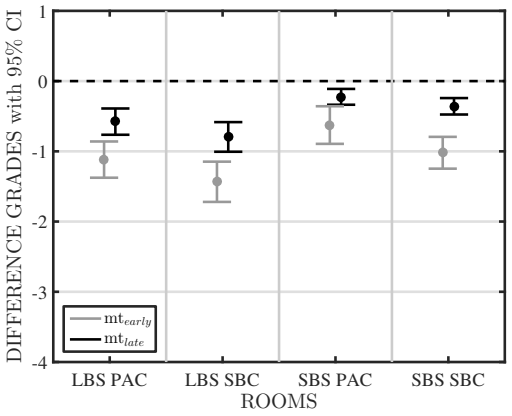

Figure 4: Mean differences grades and 95 % confidence intervals averaged over all subjects across rooms: mixing time$_{early}$ (gray), mixing time$_{late}$ (black), large room (LBS), small room (SBS), omnidirectional source (SBC), directional source (PAC)

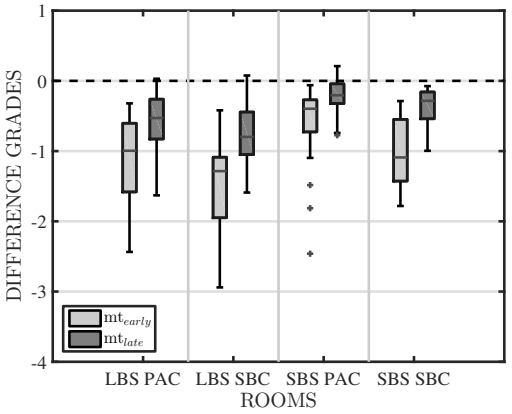

Figure 5: Difference grades of all subjects across rooms in form of a boxplot: mixing time$_{early}$ (light gray), mixing time$_{late}$ (dark grey), large room (LBS), small room (SBS), omnidirectional source (SBC), directional source (PAC)

Conclusion

A method for the synthesis of the late reverberation part of BRIRs based on a parametric model was proposed. The approach takes advantage of the perceptual mixing time and especially for dynamic binaural synthesis the reduction of the BRIR datasets is large. Although the reverberation tail is reduced to three features only, the physical evaluation shows only minor differences between original and synthesis. Furthermore the impairments caused by the synthetic reverberation were evaluated in a listening experiment. The subjects were able to identify the synthesis but we observed only small distinctions to the original BRIRs. So in general the presented approach is working satisfactorily. It is noticeable that the ratings between the different mixing times deviate each other only slightly. Considering the huge modification of the impulse responses, the ratings of the early mixing times were unexpected. Thus there is a trade-off between data volume and similarity. Depending on the used application, early mixing times could be sufficient. Probably the benefit of data reduction is greater than the auralization enhancement with later mixing times. Furthermore the influences of different settings of the approach have been perceptually evaluated in an additional study and submitted for publication in summer 2016.

Acknowledgment

The research activities are funded by the Federal Ministry of Education and Research (BMBF) in Germany under the support code 03FH005I3-MoNRa. We appreciate the great support.

References

[1] Scheirer, E. D., Vaananen, R., and Huopaniemi, J., "AudioBIFS: Describing audio scenes with the MPEG-4 multimedia standard," *IEEE Transactions on Multimedia*, 1(3), pp. 237–250, 1999.

[2] Breebaart, J., Nater, F., and Kohlrausch, A., "Parametric binaural synthesis: Background, applications and standards," *NAG/DAGA International Conference on Acoustics*, 2009.

[3] Lindau, A., Kosanke, L., and 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), pp. 887–898, 2012.

[4] Stade, P., "Perzeptive Untersuchung zur Mixing Time und deren Einfluss auf die Auralisation," in *Fortschritte der Akustik - DAGA 2015*, pp. 1103–1106, Nürnberg, 2015.

[5] Lundeby, A., Vigran, T. E., Bietz, H., and Vorländer, M., "Uncertainties of measurements in room acoustics," *Acta Acustica united with Acustica*, 81(4), pp. 344–355, 1995.

[6] Menzer, F. and Faller, C., "Investigations on an Early-Reflection-Free Model for BRIRs*," *Journal of the Audio Engineering Society*, 58(9), pp. 709–723, 2010.

[7] Moorer, J. A., "About this reverberation business," *Computer music journal*, 3(2), pp. 13–28, 1979.

[8] Menzer, F. and Faller, C., "Binaural Reverberation Using a Modified Jot Reverberator with Frequency-Dependent Interaural Coherence Matching," *Audio Engineering Society Convention 126*, 2009.

[9] Stade, P., Bernschütz, B., and Rühl, M., "A Spatial Audio Impulse Response Compilation Captured at the WDR Broadcast Studios," in *Proc. of the VDT International Convention 2012*, Cologne, 2012.

[10] ITU-R BS.1116-3, "Methods for the subjective assessment of small impairments in audio systems," 2015.

[11] Lindau, A. and Brinkmann, F., "Perceptual Evaluation of Headphone Compensation in Binaural Synthesis Based on Non-Individual Recordings," *Journal of the Audio Engineering Society*, 60(1/2), pp. 54–62, 2012.