Synthetic Reflections for Binaural Rendering using Sound Field Analysis

Philipp Stade[1,2], Johannes M. Arend[1,2]
[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

In the field of virtual audio, measured binaural room impulse responses (BRIRs) are used for the auralization of acoustical environments applying dynamic binaural synthesis. Depending on the application, it can be desirable to scale down the resolution of the BRIRs and thus reduce computational effort and the amount of data. Spherical microphone arrays can be used for the directional analysis of acoustical environments using Fourier acoustics [1]. The so-called Sound Field Analysis (SFA) is based on the same mathematical principles as the Wave Field Synthesis (WFA) spatial audio reproduction method [2] which relates the sound pressure inside a source free volume with the pressure and velocity on its surface by use of the Kirchhoff-Helmholz-Integral. It is possible to exchange the definition of inside and outside, which allows for the calculation of the sound field beyond the surface. In this paper, a new approach for the synthesis of BRIRs is presented: Combining spherical microphone array measurements with sound field decomposition techniques, it is possible to identify reflections in the measured data and to compute directional room impulse responses (DIRs) for arbitrary directions [3][4]. This enables a parametric description of the acoustical environment. Combining spherical head related impulse responses (HRIRs) with this description, synthetic BRIRs are generated which can be used for auralization purposes. The underlying parametric model and the signal processing are explained and BRIRs based on synthetic reflections are compared to the measured counterparts.

System Overview

The general idea of the presented approach is to determine the main features of measured acoustical environments and characterize them with a few parameters (see Figure 1, analysis-part). As input signal, the approach is working on impulse responses (IRs) captured with a spherical microphone array. Based on these parameters, a BRIR dataset is synthesized which can be used for auralization with dynamic binaural synthesis (see Figure 1, synthesis-part). Directional and diffuse components are processed independently from each other within the model. Direct sound and early reflections are characterized with the four directional parameters *time*, *direction*, *level* and *spectrum*. Diffuse reverberation is reduced to the two frequency-dependent diffuse parameters *energy decay curve* and *interaural coherence*.

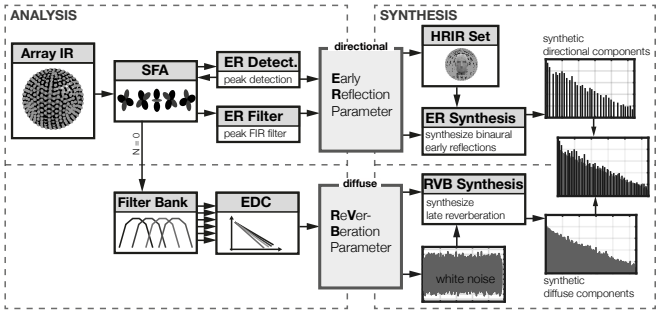

Figure 1: System Overview: Directional parameters are extracted from array measurements with sound field analysis (SFA) and used to synthesize reflections with spherical HRIRs. Diffuse reverberation parameters are determined and are used to shape white noise as a decaying masking signal.

Directional Analysis

The determination of the directional components is based on SFA [5] using the SOFiA toolbox [6]. In a first step, a spatio-temporal intensity matrix of the sound field is calculated (see Figure 2). The array IRs are temporally segmented and each time slice is transformed from time domain into the spherical wave spectrum domain using first a conventional Fourier transform and afterwards a spatial Fourier transform. In a last step, multiple plane wave decompositions are applied using the spatial Fourier coefficients, a spherical composite grid and modal radial filters. This processing depends on the chosen spherical decomposition order N and the frequency f and delivers the intensities of the sound field for each grid node and each time segment. A peak-detection algorithm is used to identify maxima (which refer to occurring reflections) in this multi-dimensional matrix (see Figure 3). The accuracy of this processing depends on the length of the time segments (temporal resolution) and the density of the used composite grid (spatial resolution).
The algorithm is summarized in the following: First, a potential maximum has to exceed a certain threshold, which refers to the mean intensity level within a time slice (*sensitivity*). After that, this maximum candidate is compared to the intensities of neighbouring nodes of the used grid within a chosen radius (*surface range*). In this step, the algorithm estimates if neighbouring intensities refer to the same reflection or not. It has to be considered that due to the used order N, the directional gain and the spatial resolution of the array is limited. Therefore a plane wave impact leads to a blurred point in the spatial response. The intensity matrix can also be spread over time, so that a number of time slices around

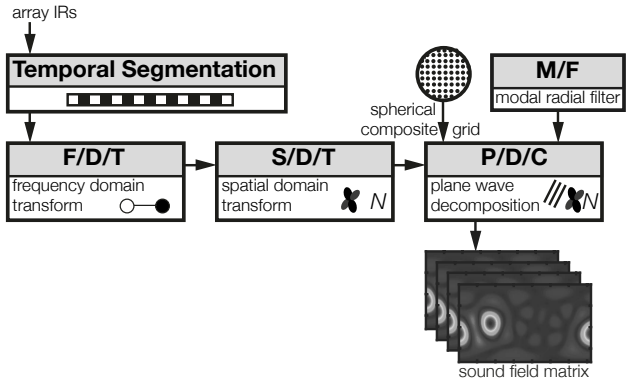

Figure 2: Calculation of the spatio-temporal intensity matrix using plane wave decomposition

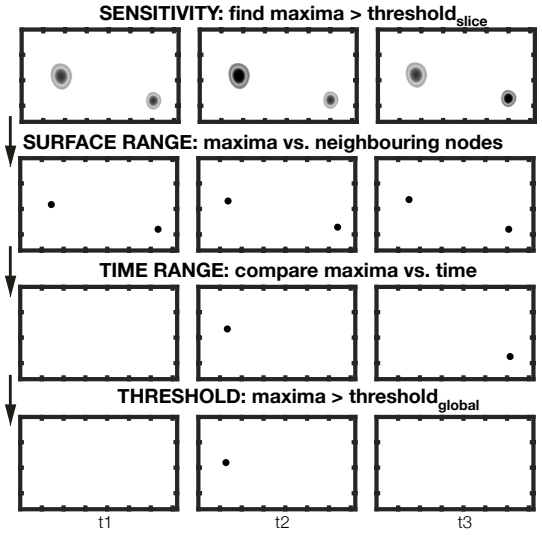

Figure 3: Exemplary presentation of the peak detection in three time slices: Potential maxima are located in the matrix and classified; one reflection identified

the slice of the potential maximum are compared to each other (*time range*). By this, the algorithm determines if the same reflection is detected in different slices or if multiple reflections occur in quick succession. Finally, if all these tests are valid and if the maximum candidate exceeds an absolute *threshold* (referring to the energy decay curve (EDC) of the normalized center impulse response of the array), the maximum is identified as a reflection and stored with its parameters in a table (*peak list*).
The previous processing calculates for each detected reflection its direction of incidence, level and delay. In addition, every reflection has a certain spectral structure due to the materials and the absorption in the room. Furthermore the audibility of reflections depends on their spectral shape [7]. Therefore DIRs are generated using again plane wave decomposition (see Figure 4) to determine the spectrum of every detected reflection. This processing depends on the entries of the *peak list* and is explained in the following for one reflection only. The array IRs are temporal segmented using a window of 192 samples around the detected *time* of the reflection. This segment is transformed again into the spherical wave spectrum domain as explained before. Then a plane wave

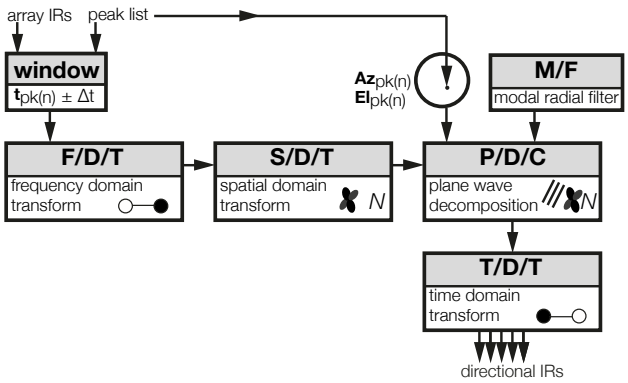

Figure 4: Generation of the directional impulse responses based on the detected peaks

decomposition with order $N = 5$ is applied in direction of the reflection using the appropriated spatial Fourier coefficients of the corresponding grid node. With inverse Fourier transform, the wave field is transferred back into the time domain and DIRs in direction of incidence of the reflection are generated. Hann-windowed linear phase FIR filters are calculated based on the DIRs to adapt the spectrum of each reflection. Linear phase filters were used to avoid spectral cancellations in the synthesis due to different phase responses of the reflection filters. The synthesis of the direct sound distinguishes slightly: The entry of the *peak list* with the maximum level is regarded as the direct sound and its spectrum is determined with an DIR with decomposition order $N = 0$ which is equal to an omnidirectional IR. Furthermore no temporal segmentation is applied. A hann-windowed minimum phase FIR filter is used for spectral adaptation of the direct sound to avoid pre-ringing artifacts at low frequencies.

Directional Synthesis

Direct sound and early reflections are synthesized using the directional parameters (*peak list* and *spectrum)* and a spherical HRIR dataset [8]. In this investigation a *Neumann KU100* has been used [9], but datasets from other artificial heads or even individual HRIR measurements can be easily integrated into the model. The HRIR dataset has to be transformed and stored in the spherical wave spectrum domain using a spatial Fourier transform. The synthesis is explained in the following for one reflection only, but the processing is applied for every entry of the *peak list* and repeated for azimuthal steps of 1° to allow for an adequate auralization considering head movements on the full horizontal plane. Other resolutions or even spherical tracking grids are possible but not applied in this study. The spatial Fourier coefficients of the HRIR dataset at the respective reflection angle (azimuth and elevation of *direction* parameter) are transformed into frequency domain with inverse spatial Fourier transform. The spectrum is adapted using the appropriated FIR filter with delay compensation. Then the signals are transformed back into time domain with inverse Fourier transformation. Depending on the *level*- and *time*-parameters, each reflection is attenuated and

delayed. Finally all signals are summed up and the output is a dataset with binaural synthetic reflections, which can be used for dynamic binaural synthesis.

Diffuse Reverberation

Beside the directional components, synthetic diffuse reverberation is generated and added to the model. Otherwise the echogram could be audible and the synthetic reflections could be perceived as delays. Therefore, a decaying noise signal is used to mask the silent frames between the synthetic reflections. The processing is based on an approach which has been proposed and perceptually evaluated in [10]. Two parameters are determined in the diffuse analysis: The *energy decay curve (EDC)* of the omnidirectional array impulse response (DIR with order $N = 0$) is calculated frequency-dependently and approximated with a polynomial curve fitting algorithm. Furthermore, the theoretical *interaural coherence (IC)* for a perfectly diffuse sound field [11] is calculated and approximated again with a polynomial. If a reference BRIR is avaliable, the accurate *IC* can be calculated and characterized with parameters as well [10]. The diffuse parameters are used to shape a dual channel white noise signal according to the envelope of the reference and to adapt its interaural coherence. First, the noise signal is processed with a polyphase filter bank with near perfect reconstruction. Each noise band is element-wise multiplied with the corresponding *EDC* curve, to shape the noise to the reference's reverberation time and spectrum. Specific filters [12] are used to adapt the uncorrelated noise signals to the reference *IC*. Finally all bands are summed up and added to the synthetic reflections.

Technical Evaluation

Two different rooms, located at the WDR Cologne, have been used for technical evaluation of the presented approach (*small broadcast studio*: $V = 1247\,m^3$, $rt60_{mean} = 0.83\,s$ and *large broadcast studio*: $V = 6098\,m^3$, $rt60_{mean} = 1.46\,s$). In both rooms, BRIR datasets with 1° resolution on the azimuthal plane (*Neumann KU100* artificial head) as a reference and spherical microphone array datasets (*VariSphear* array, 1202 sample points on Lebedev grid, diameter = 0.175 m) have been measured [13]. The reflection synthesis is based on the microphone array measurements and the previously presented approach with sound field analysis techniques. Two different gradations of the model have been generated: A synthesis using directional components only as well as a synthesis based on the entire parametric model (directional+diffuse components) are compared to the reference BRIRs in the following. In the *small broadcast studio* (see Figure 5) the first 160 ms of the impulse response using approximately 50 reflections are synthesized. Due to the greater reverberation time, the synthetic reflections of the *large broadcast studio* (see Figure 6) are based on the first 320 ms and approximately 200 reflections. The direct sound is synthezised in the same way based on the reflection with maximum level. As late reverberation

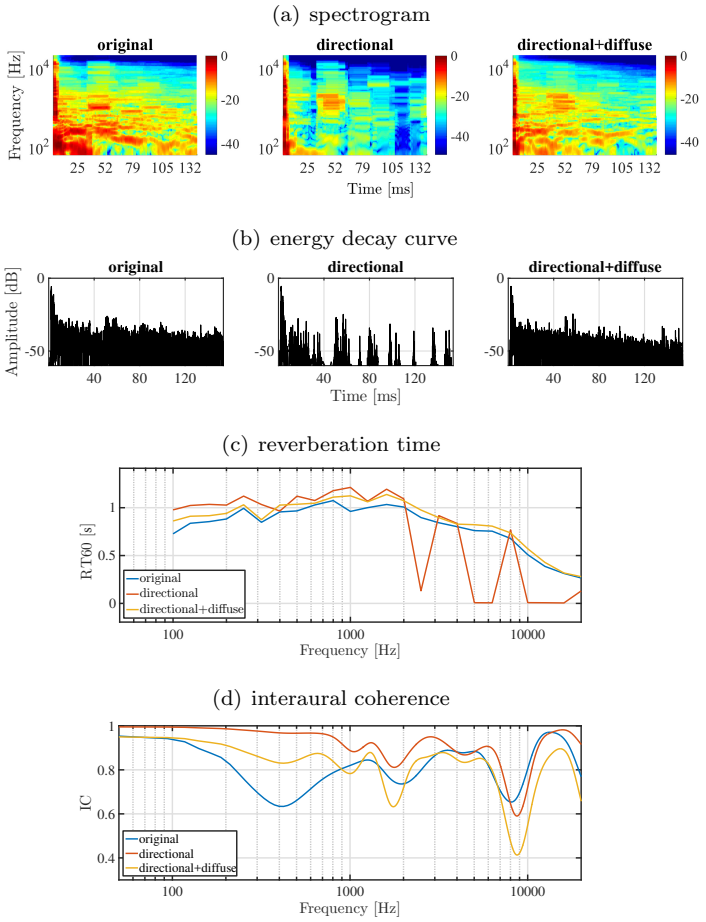

Figure 5: Small broadcast studio: spectrogram (a), energy decay curve (b), reverberation time (c) and interaural coherence (d) of original BRIR (left) vs. synthetic BRIR with directional components (middle) vs. synthetic BRIR with directional and diffuse components (right)

part, the original measured BRIR after 160 / 320 ms has been used. We observed similar results for the approach in both rooms, so that the following discussion can be transferred to all investigated scenarios. Using only the directional components of the proposed model, synthetic BRIRs with strong deviations compared to the original BRIRs were generated. The spectrograms as well the energy decay curves reveal missing energy due to silent frames in the synthetic IRs. Furthermore in some frequency bands, the reverberation time calculation fails due to the gaps in the IR and no trustful values could be determined. The interaural coherence of the synthetic BRIRs deviate widely compared to the reference. This indicates that the approach based only on a few reflections generates probably non adequate synthetic BRIR datatsets which might impair the auralization. However, using synthetic diffuse reverberation as a masking signal in addition to the synthetic reflections (directional+diffuse components), less differences between reference and synthesis were found. The spectrograms as well the energy decay curves match well and no fragmentation of the impulse responses is observed. The reverberation time tends to be only a bit longer in the synthesis and the matching of the interaural coherence

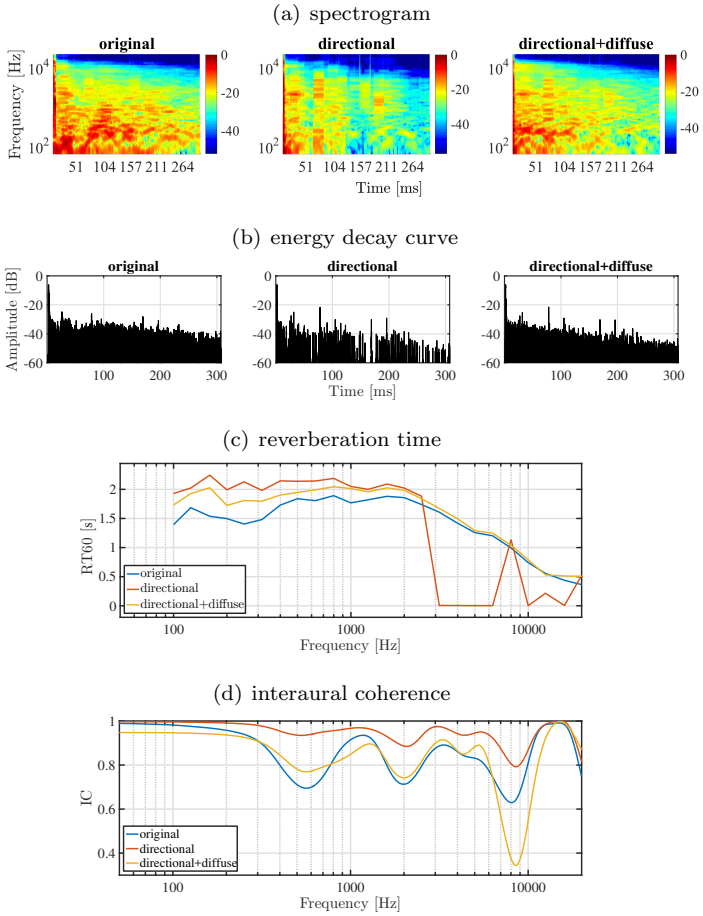

Figure 6: Large broadcast studio: spectrogram (a), energy decay curve (b), reverberation time (c) and interaural coherence (d) of original BRIR (left) vs. synthetic BRIR with directional components (middle) vs. synthetic BRIR with directional and diffuse components (right)

is more precisely than when using just reflections. Only in some frequency bands, bigger variations were still found. But in general the synthesis using directional and diffuse components works quite accurate.

Conclusion

A parametric model for the synthesis of BRIRs using sound field analysis was presented. The approach is based on microphone array measurements and focuses on a plausible auralization with less computational effort. Ideally the reduced resolution of the synthetic BRIRs is not perceptible by the recipient. The array data is analyzed and an echogram with the direction, time, level and spectrum of each detected reflection is determined with plane wave decomposition methods. Furthermore, diffuse components like the frequency-dependently EDC and the IC are calculated. The synthesis is based on these parameters only using a previously captured spherical HRIR dataset. Synthetic BRIRs based on this model as well as synthetic BRIRs using only synthetic reflections have been technically examined and compared to measured BRIRs. While the synthesis with directional parts only shows a non adequate performance, the entire model works quite satisfying with less variances as against the reference. This demonstrates the importance of the masking synthetic diffuse reverberation even in the early part of a BRIR, especially if a highly simplified echogram is used. In future, the approach and its performance in different rooms has to be evaluated perceptually. Results of the corresponding listening experiment will be presented in [14].

Acknowledgement

This work was funded by the Federal Ministry of Education and Research (BMBF) under the support code 03FH005I3-MoNRa. We appreciate the great support.

References

[1] Williams, E. G., *Fourier Acoustics*, Sound Radiation and Nearfield Acoustical Holography, Academic Press, 1999.

[2] Spors, S., Rabenstein, R., and Ahrens, J., "The theory of wave field synthesis revisited," *Audio Engineering Society Convention 126*, 2008.

[3] Meyer, J. and Elko, G., "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," *Proc. of ICASSP '02*, 2, pp. II–1781–II–1784, 2002.

[4] Duraiswami, R., Li, Z., Grassi, E., Gumerov, N. A., and Zotkin, D., "Plane-wave decomposition analysis for spherical microphone arrays," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 150–153, 2005.

[5] Bernschütz, B., Stade, P., and Rühl, M., "Sound Field Analysis in Room Acoustics ," in *Proc. of the VDT International Convention 2012*, Cologne, 2012.

[6] Bernschütz, B., Pörschmann, C., Spors, S., and Weinzierl, S., "SOFiA sound field analysis toolbox," in *Proc. of the International Conference on Spatial Audio - ICSA 2011*, Detmold, 2011.

[7] Olive, S. E. and Toole, F. E., "The detection of reflections in typical rooms," *Journal of the Audio Engineering Society*, 37(7/8), pp. 539–553, 1989.

[8] Duraiswami, R., Zotkin, D., Grassi, E., Gumerov, N. A., and Davis, L. S., "High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues," in *Proc. of the 119th AES Convention*, New York, 2005.

[9] Bernschütz, B., "A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100," in *AIA DAGA 2013*, pp. 592–595, Meran, 2013.

[10] Stade, P. and Arend, J. M., "Perceptual Evaluation of Synthetic Late Binaural Reverberation Based on a Parametric Model ," in *Proc. of the AES International Conference on Headphone Technology*, pp. 80–87, Aalborg, 2016.

[11] Cook, R. K., "Measurement of Correlation Coefficients in Reverberant Sound Fields," *The Journal of the Acoustical Society of America*, 27(6), pp. 1072–1077, 1955.

[12] Menzer, F. and Faller, C., "Investigations on modeling BRIR tails with filtered and coherence-matched noise," in *Proc. of the Audio Engineering Society Convention 127*, 2009.

[13] 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.

[14] Stade, P., Arend, J. M., and Pörschmann, C., "Perceptual Evaluation of Synthetic Early Binaural Room Impulse Responses Based on a Parametric Model," in *Proc. of the 142nd AES Convention*, pp. 1–10, Berlin, 2017.