ABSTRACT
The auralization of acoustic environments over headphones is often realized with data-based dynamic binaural synthesis. The required binaural room impulse responses (BRIRs) for the convolution process can be acquired by performing measurements with an artificial head for different head orientations and positions. This procedure is rather costly and therefore not always feasible in practice. Because a plausible representation is sufficient for many practical applications, a simpler approach is of interest. In this paper we present the BinRIR (Binauralization of omnidirectional room impulse responses) algorithm, which synthesizes BRIR datasets for dynamic auralization based on a single measured omnidirectional room impulse response (RIR). Direct sound, early reflections, and diffuse reverberation are extracted from the omnidirectional RIR and are separately spatialized. Spatial information is added according to assumptions about the room geometry and on typical properties of diffuse reverberation. The early part of the RIR is described by a parametric model and can easily be modified and adapted. Thus the approach can even be enhanced by considering modifications of the listener position. The late reverberation part is synthesized using binaural noise, which is adapted to the energy decay curve of the measured RIR.

In order to examine differences between measured and synthesized BRIRs, we performed a technical evaluation for two rooms. Measured BRIRs are compared to synthesized BRIRs and thus we analyzed the inaccuracies of the proposed algorithm.

1. INTRODUCTION
Binaural synthesis is a powerful tool for headphone-based presentation of virtual acoustic environments (VAEs). It can be applied for auralization purposes in various areas like audio engineering, telecommunication, or architectural acoustics. For many of these applications, a plausible presentation is sufficient; an authenticity reproduction of the sound field is often not pursued. In this context plausibility refers to the illusion that the scenario being depicted is actually occurring [1] while authentic refers to a perception that the scenario cannot be distinguished from a real reference.

Binaural room impulse responses (BRIRs) can be applied, which are either simulated or measured with an artificial head for different orientations (and positions). Finally the BRIRs are convolved with anechoic signals in a binaural renderer. By considering the listener’s head movements in the auralization, localization accuracy increases [2]. Front-back confusion can be decreased [2] and externalization of virtual sound sources improves [3][4]. Several commercial or scientific rendering engines are available, which adapt the sound field presented through headphones according to the orientation of the listener in real time (e.g. [5][6][7][8]). Depending on the head movements, which shall be considered in the auralization, these measurements need to be done for circular orientations in the horizontal plane or even for spherical head orientations considering horizontal and vertical rotations. However, measuring such BRIR datasets requires a large amount of time and the use of complex devices (e.g. a rotatable artificial head). Furthermore, for each listening position, another set of BRIRs needs to be captured. Thus, for many applications in the field of spatial audio and virtual environments, the effort is so high that circular sets of BRIRs are not used. To approach this issue, we developed the BinRIR (Binauralization of omnidirectional room impulse responses) algorithm, which aims for an auralization based on a simple measurement procedure. Only one single measured omnidirectional room impulse response (RIR) is required to obtain a plausible auralization when using dynamic binaural synthesis. The algorithm even allows to shift the listener position. Thus, one single measured RIR is sufficient to synthesize a BRIR dataset for a freely chosen head orientation and position in the room.

In literature, several approaches to obtain BRIRs from measured RIRs have been described. In [9][10] a synthesis of BRIRs from B-format measurements has been proposed. The spatial impulse response rendering (SIRR) method applies a decomposition of the sound field into direct and diffuse parts. While the diffuse part is decorrelated, vector-based amplitude panning is used to distribute the direct sound on different loudspeakers. In [11] a directional audio coding (DirAC) method is proposed which can capture, code, and resynthesize spatial sound fields. DirAC analyzes the audio signal in short time frames and determines the spectrum together with direction and diffuseness in the frequency bands of human hearing. As this method does not work impulse response-based it is quite different to the one presented in this paper. Another simple approach to synthesize BRIRs has been presented by Menzen.

In [12][13] RIRs measured in the B-format are used to synthesize BRIRs. Direct sound, spectral shape and temporal structure are extracted from the RIR. Additionally, the incidence direction of the direct sound is estimated from the measured data. No specific treatment of the early reflections is proposed. All reflections and the diffuse reverberation are synthesized by performing an adequate reconstruction of the interaural coherence. In this paper, we present research results on the binauralization of omnidirectionally measured RIRs. Parts of the studies including the basic idea and a basic description of the approach as well as results of a perceptual evaluation have already been published [14][15][16]. This paper is organized as follows: In section 2 we introduce and explain the BinRIR algorithm performing the spatialization of omnidirectional RIRs in detail. In section 3 we describe the results of a technical evaluation. We compare measured BRIRs of two different rooms to the synthesized counterparts and elaborate differences caused by the simplifications of the algorithm. Finally, section 4 concludes the paper and provides an outlook.
2. ALGORITHM DESIGN

2.1. General Structure

The basic idea of the BinRIR algorithm is to use only one measured omnidirectional RIR for the synthesis of BRIR datasets which can be used for dynamic auralization. The algorithm was implemented in Matlab and applies predictable information from sound propagation in enclosed spaces as well as knowledge regarding the human perception of diffuse sound fields. For processing, the RIR is split into different parts. The early part contains the direct sound and strong early reflections. For this part, the directions of incidence are modeled reaching the listener from arbitrarily chosen directions. The late part of the RIR is considered being diffuse and is synthesized by convolving binaural noise with small sections of the omnidirectional RIR. By this, the properties of diffuse reverberation are approximated. The algorithm includes an additional enhancement: The synthesized BRIRs can be adapted to shifts of the listener and thus freely chosen positions in the virtual room can be auralized.

The BinRIR algorithm incorporates several inaccuracies and deviates significantly from a measured BRIR. The directions of incidence of the synthesized early reflections are not in line with the real ones. Hence, differences in the perception of spatial properties (e.g. envelopment) between the original room and the synthesized room may occur. Furthermore, a point source is assumed for all synthetic BRIR datasets. Thus, it is not possible to rebuild source width and other properties of the source correctly. Finally, the diffusely reflected part of the early reflections cannot be precisely reconstructed.

The basic structure of the BinRIR algorithm is shown in Figure 1. As input data the algorithm only requires the omnidirectional RIR and the position of the sound source. Furthermore, the algorithm accesses an appropriate set of HRIRs and a preprocessed sequence of binaural noise. Both were obtained from measurements with a Neumann KU 100 artificial head [17]. The algorithm is only applied to frequencies above 200 Hz. For lower frequencies the interaural coherence of a typical BRIR is nearly one and the omnidirectional RIR can be maintained. 7th order Chebyshev Type II filters are used to separate the low frequency part from the rest of the signal.

2.2. Direct sound and early reflections

Onset detection is used to identify the direct sound in the omnidirectional RIR. The direct sound frame starts with the onset and ends after 10 ms (5 ms followed by 5 ms raised cosine offset). The following time section is assigned to the early reflections and the transition towards the diffuse reverberation. For small and non-reverberant rooms (V < 200 m$^3$ and $\text{RT}_60 < 0.5$ s) a section length of 50 ms is chosen, otherwise the section is extended to 150 ms.

In order to determine sections with strong early reflections in the omnidirectional RIR, the energy is calculated in a sliding window of 8 ms length and time sections which contain high energy are marked. Peaks which are 6 dB above the RMS level of the sliding window are determined and assigned to geometric reflections. A windowed section (raised cosine, 5 ms ramp) around each of the peaks is considered as one reflection. If several very dense reflections occur in adjacent sections, these sections are merged. Following this procedure, small windowed sections of the omnidirectional RIR are extracted describing the early reflections. The incidence directions of the synthesized reflections base on a spatial re-

![Figure 1: Block diagram of the BinRIR algorithm for synthesizing a BRIR based on one single omnidirectional RIR.](image-url)
Figure 2: Synthesis of the binaural diffuse reverberation: Sections of the diffuse omnidirectional RIR (0.67 ms; 32 taps at 48 kHz sampling rate) and the binaural noise (2.67 ms; 128 taps at 48 kHz sampling rate) are convolved. Both sections are raised-cosine windowed. The diffuse BRIR is synthesized by summing up the results of the convolutions applying overlap-add.

2.4. Listener position shifts

The algorithm includes a further enhancement: The synthesized BRIR can be adapted to listener position shifts (LPS) and thus freely chosen positions of the listener in the virtual room can be auralized. For this, a simple geometric model based on mirror-image sound sources is used. The distance between the listener and each of the mirror-image sound sources is determined from the delay of the corresponding reflection peak to the direct sound peak. In a next step, a shifted position of the listener is considered and amplitudes (based on the 1/r law), distances, and directions of incidence are recalculated for each reflection (Fig. 3). Optimizing an earlier version of the BinRIR algorithm (e.g. [16]) we modified the low-frequency component below 200 Hz when applying LPS. If, for example, the listener approaches the sound source, the amplitude of the direct sound increases and the low-frequency energy for the direct sound needs to be adapted accordingly. For this the low-frequency part of the direct sound (first 10 ms followed by 10 ms raised cosine set) is adjusted according to the 1/r law.

2.5. Synthesis of Circular BRIR sets

The synthesis of the BRIRs is repeated for constant shifts in the azimuthal angle (e.g. 1°) for the direct and the reflected sound. Thus, a circular set of BRIRs is obtained, which can be applied by different rendering engines for dynamic binaural synthesis. The synthesized sets of BRIRs can be stored in various formats, e.g. the micro-Format [24], a multi-channel-wave format to be used by the SoundScape Renderer [15] and can be converted to the SOFA-format [25].

3. TECHNICAL EVALUATION

To analyze the performance of the algorithm, a series of listening experiments has already been conducted [14][16]. These experiments mainly aimed at a quantification of the perceptual differences between measured and synthesized BRIRs. In this paper we focus on a technical evaluation of the algorithm and compare different properties of the synthesized BRIRs to the properties of measured BRIRs. Therefore we analyzed the detected reflections regarding their direction, their time of incidence, and their amplitude to measured data. Furthermore, we compared the reverberation tails and the reverberation times of the synthesized BRIRs to the ones of measured BRIRs. We investigated to what extent the clarity (C50) of the synthesized room matches the measured room’s
Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17), Edinburgh, UK, September 5–9, 2017

3.1. Measured rooms

The performance of the algorithm was analyzed for two different rooms. Both rooms are located at the WDR radiobroadcast studio in Cologne and are used for various recordings of concerts and performances. The “KVB-Saal” (Large Broadcast Studio - LBS) has a volume of 6100 m$^3$, a base area of 579 m$^2$ and can seat up to 637 persons. We measured the impulse responses in the 6th row (Distance$_{SrcRec}$ = 13.0 m). The “kleiner Sendesaal” (Small Broadcast Studio - SBS) has a volume of 1247 m$^3$, a base area of 220 m$^2$ and 160 seats. The Distance$_{SrcRec}$ in this room was 7.0 m. In order to evaluate the algorithm, measured impulse responses from these rooms were used. In addition to the omnidirectional RIRs, which are required to feed the BinRIR algorithm, we measured circular BRIR datasets at the same position as a reference. This dataset was measured in steps of 1° on the horizontal plane with a Neumann KU100 artificial head. Finally we used data from spherical microphone array measurements, conducted with the VariSphear measurement system [26]. For this, we applied a rigid sphere array configuration with 1202 sample points on a Lebedev grid at a diameter of 17.5 cm. The omnidirectional RIRs and the array data were measured with an Earthworks M30 microphone. As sound source, a PA stack involving an AD Systems Stium Mid/High unit combined with 3 AD Systems Flex 15 subwoofers was used. The complete series of measurements is described in detail in [27].

Based on the microphone array measurements, we identified reflections in the room using sound field analysis techniques [28]. For this the array impulse responses were temporally segmented (time resolution = 0.5 ms) and transformed into the frequency domain. Applying a spatial Fourier transformation, the impulse responses were transformed into the spherical wave spectrum domain [29]. Then the sound field was decomposed into multiple plane waves using the respective spatial Fourier coefficients. Data was extracted for a decomposition order of N = 5, a spherical composite grid with 3074 Lebedev points at a frequency $f = 4500$ Hz. For this frequency quite robust results for the detection of the reflections were found. By this, a spatio-temporal intensity matrix of the sound field at the listener position was calculated. Each time slice of the matrix was analyzed with a specific algorithm in order to classify reflections which are represented as local maxima in the matrix [30]. The detected reflections were stored with their attributes “time”, ”direction” and ”level” in a reflection list and can be used for further comparison with the BinRIR algorithm.

3.2. Direct Sound and early reflections

In a first step we looked at the early part of the synthetic BRIRs and compared the direct sound and the early reflections determined by the BinRIR algorithm to the reflections which are identified using sound field analysis techniques based on the microphone array measurements. To describe the directions of the reflections we used a spherical coordinate system. The azimuthal angle denotes the orientation on the horizontal plane with $\varphi = 0^\circ$ corresponding to the front direction. The elevation angle is $\delta = 0^\circ$ for frontal orientation and $\delta = 90^\circ$ for sound incidence from above.

Table 1 and 2 show the reflections detected by BinRIR as well as the 14 strongest reflections which were identified based on the array data. For each reflection, the time of arrival relative to the direct sound, the incidence direction, and the level are shown. The temporal structure and the energy of many of the reflections show similarities. For example in the LBS for reflection #7, #10 and #13 and in the SBS for reflection #13 the level and the time of arrival match quite well. Furthermore, some of the reflections were detected in the array measurements as impinging from different directions nearly simultaneously. These reflections are typically merged to one reflection with an increased level by the BinRIR algorithm based on a lookup-table, it is not surprising that there are significant differences compared to the directions of the measured reflections. However, if the proportions as well as source and listener position of the synthesized room to some extent match the modelled shoebox room some of the incidence directions are appropriate. Thus, at least for the first reflection and as well for some other reflections detected by BinRIR, an acceptable congruence exists both for the LBS and the SBS. The azimuthal deviation of the first reflection is less than 40° and in elevation no relevant deviation exists.

As already explained in 2.2 the BinRIR algorithm starts detecting reflections 10 ms after the direct sound. Thus no reflections reaching the listener within the first 10 ms can be found. The time frame

![Figure 4](image-url)

Figure 4: Reverberation Time ($RT_{60}$) of the Large Broadcast Studio (a) and the Small Broadcast Studio (b). In each plot the $RT_{60}$ for the binaurally measured reference, for the synthesis with the BinRIR algorithm and for the omnidirectional measurement are shown. The $RT_{60}$ was calculated in 1/3 octave bands in the time domain.

clarity. Finally we looked briefly in which way the use of the LPS influences the early part of the BRIR.
BinRIR algorithm

<table>
<thead>
<tr>
<th>#</th>
<th>Delay [ms]</th>
<th>Azimuth [°]</th>
<th>Elevation [°]</th>
<th>Level [dB]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>5.0</td>
<td>8</td>
<td>-58</td>
<td>-27.6</td>
</tr>
<tr>
<td>2</td>
<td>7.5</td>
<td>142</td>
<td>-18</td>
<td>-27.5</td>
</tr>
<tr>
<td>3</td>
<td>17.5</td>
<td>8</td>
<td>8</td>
<td>-27.3</td>
</tr>
<tr>
<td>4</td>
<td>24.0</td>
<td>309</td>
<td>0</td>
<td>-25.6</td>
</tr>
<tr>
<td>5</td>
<td>26.5</td>
<td>58</td>
<td>5</td>
<td>-27.9</td>
</tr>
<tr>
<td>6</td>
<td>28.0</td>
<td>278</td>
<td>0</td>
<td>-28.1</td>
</tr>
<tr>
<td>7</td>
<td>31.0</td>
<td>0</td>
<td>58</td>
<td>-20.3</td>
</tr>
<tr>
<td>8</td>
<td>38.0</td>
<td>253</td>
<td>-2</td>
<td>-28.0</td>
</tr>
<tr>
<td>9</td>
<td>50.5</td>
<td>180</td>
<td>20</td>
<td>-27.2</td>
</tr>
<tr>
<td>10</td>
<td>79.5</td>
<td>178</td>
<td>17</td>
<td>-12.7</td>
</tr>
<tr>
<td>11</td>
<td>92.5</td>
<td>353</td>
<td>73</td>
<td>-26.6</td>
</tr>
<tr>
<td>12</td>
<td>108.2</td>
<td>0</td>
<td>0</td>
<td>-31.4</td>
</tr>
<tr>
<td>13</td>
<td>117.6</td>
<td>85</td>
<td>0</td>
<td>-31.4</td>
</tr>
<tr>
<td>14</td>
<td>174.0</td>
<td>356</td>
<td>8</td>
<td>-25.0</td>
</tr>
<tr>
<td>15</td>
<td>202.0</td>
<td>3</td>
<td>3</td>
<td>-24.3</td>
</tr>
</tbody>
</table>

Table 1: Properties of the direct sound and the early reflections for the Large Broadcast Studio (LBS). Left side: Reflections determined from the analysis of the array data. Right side: Reflections detected by the BinRIR algorithm.

<table>
<thead>
<tr>
<th>#</th>
<th>Delay [ms]</th>
<th>Azimuth [°]</th>
<th>Elevation [°]</th>
<th>Level [dB]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>1.5</td>
<td>7</td>
<td>-22</td>
<td>-6.9</td>
</tr>
<tr>
<td>2</td>
<td>3.5</td>
<td>16</td>
<td>-76</td>
<td>-21.9</td>
</tr>
<tr>
<td>3</td>
<td>7.5</td>
<td>210</td>
<td>-22</td>
<td>-27.5</td>
</tr>
<tr>
<td>4</td>
<td>12.0</td>
<td>204</td>
<td>-17</td>
<td>-27.7</td>
</tr>
<tr>
<td>5</td>
<td>15.5</td>
<td>27</td>
<td>66</td>
<td>-23.9</td>
</tr>
<tr>
<td>6</td>
<td>20.0</td>
<td>284</td>
<td>0</td>
<td>-21.1</td>
</tr>
<tr>
<td>7</td>
<td>23.0</td>
<td>58</td>
<td>75</td>
<td>-24.5</td>
</tr>
<tr>
<td>8</td>
<td>35.0</td>
<td>203</td>
<td>2</td>
<td>-27.7</td>
</tr>
<tr>
<td>9</td>
<td>50.0</td>
<td>152</td>
<td>1</td>
<td>-21.5</td>
</tr>
<tr>
<td>10</td>
<td>50.5</td>
<td>151</td>
<td>-2</td>
<td>-21.5</td>
</tr>
<tr>
<td>11</td>
<td>51.5</td>
<td>147</td>
<td>29</td>
<td>-25.5</td>
</tr>
<tr>
<td>12</td>
<td>57.0</td>
<td>129</td>
<td>0</td>
<td>-21.4</td>
</tr>
<tr>
<td>13</td>
<td>59.0</td>
<td>173</td>
<td>7</td>
<td>-19.7</td>
</tr>
<tr>
<td>14</td>
<td>99.5</td>
<td>331</td>
<td>-2</td>
<td>-19.5</td>
</tr>
</tbody>
</table>

Table 2: Properties of the direct sound and the early reflections for the Small Broadcast Studio (SBS). Left side: Reflections determined from the analysis of the array data. Right side: Reflections detected by the BinRIR algorithm.

determining the geometric reflections ends after 150 ms and no reflections are detected by BinRIR after this period. Furthermore, several reflections which can be extracted from the array data measurements are not detected by BinRIR (e.g. # 3 and # 9 in the LBS and # 5 and # 12 in the SBS). However, in total more than 2/3 of the reflections in the section from 10 ms to 150 ms determined from the array data correspond to a reflection determined by BinRIR.

3.3. Energy Decay and Reverberation Time

In a next step we compared the energy decay curves of the synthesized BRIRs and of the reference BRIRs (Figure 5). We calculated the IC according to [31] applying hamming-windowed blocks with a length of 256 taps (5.33 ms) and an overlap of 128 taps (2.67 ms). In each plot the IC calculated with three different starting points is shown. For reference and synthesis in both rooms the IC is significantly different when direct sound is included in the calculation (t > 0 ms). For the medium condition (LBS: t > 150 ms; SBS: t > 50 ms) significant differences between synthesis and reference can be observed. This is not surprising as no shaping of the IC is included in the BinRIR algorithm. However, this difference is smaller for the SBS, because the impulse response is probably nearly diffuse at 50 ms. For the condition with the maximal starting point (LBS: t > 300 ms; SBS: t > 150 ms) which mainly comprises the diffuse reverberation, the IC of the synthesized BRIR matches the reference quite well.

3.4. Interaural Coherence

Next, we compared the interaural coherence (IC) of the synthesized BRIRs and of the reference BRIRs (Figure 5). We calculated the IC according to [31] applying hamming-windowed blocks with a length of 256 taps (5.33 ms) and an overlap of 128 taps (2.67 ms). In each plot the IC calculated with three different starting points is shown. For reference and synthesis in both rooms the IC is significantly different when direct sound is included in the calculation (t > 0 ms). For the medium condition (LBS: t > 150 ms; SBS: t > 50 ms) significant differences between synthesis and reference can be observed. This is not surprising as no shaping of the IC is included in the BinRIR algorithm. However, this difference is smaller for the SBS, because the impulse response is probably nearly diffuse after 50 ms.
Figure 5: Interaural coherence (IC) of the Large Broadcast Studio (LBS) and the Small Broadcast Studio (SBS). In plot (a) and (b) the data for the binaural reference is shown, in (c) and (d) the data for the BinRIR synthesis. For the LBS the IC is plotted for $t > 0\,\text{ms}$, $t > 150\,\text{ms}$ and $t > 300\,\text{ms}$, for the SBS for $t > 0\,\text{ms}$, $t > 50\,\text{ms}$ and $t > 150\,\text{ms}$.

Figure 6: Clarity ($C_{50}$) of the Large Broadcast Studio (LBS) and the Small Broadcast Studio (SBS). In each plot the $C_{50}$ for the binaurally measured reference, for the synthesis with the BinRIR algorithm and for the omnidirectional measurement are shown.
3.6. Listener position shifts

Finally we analyzed for the SBS in which way the early part of the synthesized BRIR is changed when listener position shifts (LPS) are performed. The results are shown in Figure 7 for synthesized Distance Factors (DFs) from 0.125 to 2 are shown in different colors.

Figure 7: Influence of the Listener Position Shift (LPS) on the early part of the time response for the Small Broadcast Studio (SBS). The time responses for Distance Factors (DFs) from 0.125 to 2 are shown in different colors.

3.5. Clarity

Next, we examined the clarity C50 over frequency for each of the conditions (Figure 6). The differences between the omnidirectional RIR, the synthesis and the reference are minor. The average unsigned deviation between the synthesis and the reference is 1.4 dB for the LBS and 1.1 dB for the SBS. The maxima are 5.2 dB (LBS) and 3.2 dB (SBS). Thus the ratio of the energy of the early part of the BRIR and the late diffuse part of the BRIR can be regarded as appropriate.

6. REFERENCES


Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17), Edinburgh, UK, September 5–9, 2017


