**Introduction**

The confusion of front and back directions is a typical problem in static binaural rendering due to ambiguous interaural cues and thus solely monaural spectral differences. These differences reduce towards the interaural axis, i.e. close to ±90°, causing more confusions at lateral directions [1]. The confusion rate can be reduced by the application of individual head-related impulse responses (HRIRs) due to slight asymmetries [2]. Besides the individual measurement of HRIRs, they can be individualized based on anthropometric data [3]. Individualization can also be achieved by additional drivers in the headphones around the pinna that activate the individual directional cues [4, 5, 6].

The confusion can further be resolved by dynamic rendering that incorporates movements of the source or the listener [7, 8, 9, 10]. In practice, movement is typically limited to tracking the orientation of the listener's head, as in [11, 12, 13].

However, in many applications, binaural rendering can neither use individual HRIRs nor high-quality head tracking. Nevertheless, experience with typical 360° video players reveals unexpectedly good discrimination between front and back directions, suggesting some modifications of the binaural signal during playback.

This contribution presents an approach to reduce front-back confusions in static binaural rendering that is weaker compared to the modifications applied by the above-mentioned 360° video players and can be used in any application that is based on HRIRs by simply exchanging the original HRIRs by modified HRIRs. Subsequent to the derivation of the approach from spectral cues that are already present in HRIRs, a listening experiment evaluates its effectiveness for discrete and first-order Ambisonics rendering.

**Approach**

The first step towards minimization of front-back confusions is a spectral analysis of HRIR pairs that are symmetrically arranged around the interaural axis, i.e. directions on the same cone of confusion.

![Figure 1: Facebook 360 Spatial Workstation Control Plugin.](image)

The control plugin of the Facebook 360 Spatial Workstation provides an option to define a focus area of adjustable size that follows the listener’s gaze, cf. Figure 1. All directions outside the focus area are attenuated by an adjustable level. Thus, in the example from Figure 1, back directions are attenuated by 24 dB.

Similarly, Youtube gradually attenuates non-frontal directions. At the back, the attenuation was measured to be about 6 dB, cf. Figure 2. More elaborated, VLC media player additionally applies a direction-dependent high-shelf filter at approx. 700 Hz for directions behind the interaural axis. The attenuation of the high-shelf reaches its maximum value of about 7 dB at the back. This follows the findings in [14] that low-pass filtered stimuli are consistently localized from the back.

![Figure 2: Frequency-dependent attenuation of sound from the back in relation to frontal sound, measured for Youtube and VLC media player.](image)

![Figure 3: Spectral differences between HRIRs (dataset from [15]) of symmetrical directions in the horizontal plane.](image)
Figure 3 shows that all pairs from $0^\circ$/$180^\circ$ to $50^\circ$/$130^\circ$ share the similar shape of a high-shelf at 3kHz. While the peak at 8kHz increases for more lateral directions, the low-pass effect above 10kHz becomes weaker. The differences clearly vanish for the pairs closer to the interaural axis. A simple model of the spectral behavior can be a 3kHz high-shelf filter with a gain of -6dB for directions behind $\pm 130^\circ$ that reduces its gain to 0dB at 90$^\circ$. This filter avoids additional cues for height in the 8kHz region [16] and partly keeps the low-pass above 10kHz. Our assumption is that applying this filter to HRIRs exaggerates the already present spectral cues and thus reduces front-back confusions, cf. Figure 4. In order to further emphasize the directional bands for the front [16], the filter gain is shifted, so that high frequencies are boosted by 3dB at the front and attenuated by 3dB at the back, cf. gray curve for $g_{max} = 3$dB in Figure 5.

\[ g_{\phi} \triangleq \begin{cases} \text{3kHz High-Shelf} & \text{if } \phi \neq \phi_L, \phi_R \\ 0 & \text{else} \end{cases} \]

**Figure 4:** Direction-dependent filter for HRIRs to reduce front-back confusions.

The effectiveness of the presented approach is evaluated in a listening experiment using HRIRs or BRIRs (binaural room impulse responses, BRIRs) from a first-order 2D image-source model, cf. Figure 7. The simulated room had a size of $5m \times 8m$ and a frequency-independent reflection factor of 0.5 with the virtual loudspeakers arranged on a radius of $2m$ around a listener at $(-0.2, 0.2)m$. The playback employed equalized AKG K-702 headphones driven by the headphone amplifier of an RME babyface in a studio with visible loudspeakers to support externalization.

**Table 1:** Binaural rendering conditions in the experiment.

<table>
<thead>
<tr>
<th>Renderer</th>
<th>IRs</th>
<th>room</th>
<th>$g_{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>discrete</td>
<td>HRIR</td>
<td>no</td>
<td>0 dB</td>
</tr>
<tr>
<td>discrete</td>
<td>HRIR$_3$</td>
<td>no</td>
<td>3 dB</td>
</tr>
<tr>
<td>discrete</td>
<td>HRIR$_6$</td>
<td>no</td>
<td>6 dB</td>
</tr>
<tr>
<td>discrete</td>
<td>BRIR</td>
<td>yes</td>
<td>0 dB</td>
</tr>
<tr>
<td>discrete</td>
<td>BRIR$_3$</td>
<td>yes</td>
<td>3 dB</td>
</tr>
<tr>
<td>discrete</td>
<td>BRIR$_6$</td>
<td>yes</td>
<td>6 dB</td>
</tr>
<tr>
<td>Ambisonics</td>
<td>HRIR</td>
<td>no</td>
<td>0 dB</td>
</tr>
<tr>
<td>Ambisonics</td>
<td>HRIR$_3$</td>
<td>no</td>
<td>3 dB</td>
</tr>
<tr>
<td>Ambisonics</td>
<td>HRIR$_6$</td>
<td>no</td>
<td>6 dB</td>
</tr>
<tr>
<td>Ambisonics</td>
<td>BRIR</td>
<td>yes</td>
<td>0 dB</td>
</tr>
<tr>
<td>Ambisonics</td>
<td>BRIR$_3$</td>
<td>yes</td>
<td>3 dB</td>
</tr>
<tr>
<td>Ambisonics</td>
<td>BRIR$_6$</td>
<td>yes</td>
<td>6 dB</td>
</tr>
</tbody>
</table>

A total of 10 listeners participated in the experiment (average age 31 years, all male). All of them were experienced in spatial audio.

**Results**

In general, listeners found the high-frequency boost of 6dB for frontal directions unnaturally sharp, while the attenuation at the back with the same gain was subtle. Although they reported the part using BRIRs to be more simple due to the better externalization, both parts took a similar time of about 9 minutes each. What is more, there was no significant difference between the average confusion rate using HRIRs or BRIRs ($p = 0.54$). This agrees with the findings in [18] that better externalization does not necessarily reduce the amount of front-back confusions. Thus, the results of both parts are summarized in the following analysis and Figures 8 and 9.
On average, the confusion rate using discrete rendering without shelving filter is 32.5%, agreeing with a value of 31% for non-individual HRIRs in [19]. The rates can be reduced to 15% for the shelving filter with ±3 dB and less than 4% for ±6 dB.

In detail, there is no significant reduction from 0 dB to 3 dB ($p = 0.25$), a weakly significant reduction from 3 dB to 6 dB ($p = 0.083$), but a significant reduction from 0 dB to 6 dB ($p = 0.007$) for the $0^\circ$ direction. For $±45^\circ$, the reduction is again weakly significant from 0 dB to 3 dB ($p = 0.084$) and significant from 0 dB to 6 dB ($p = 0.029$), although not significant from 3 dB to 6 dB ($p = 0.78$). Behind the interaural axis, the effect of the shelving filter is stronger: For $±135^\circ$, there is a weakly significant improvement from 0 dB to 3 dB ($p = 0.062$) and significant improvements from 3 dB to 6 dB, as well as from 0 dB to 6 dB ($p \leq 0.004$). All filter setting yield significant differences for $180^\circ$ ($p \leq 0.009$).

The first-order Ambisonics rendering produces on average 41.25% confusions without filtering that can be reduced to 27.5% and 15%, respectively. The worse performance in comparison to the discrete rendering is caused by coloration of the simple decoder and can be assumed to improve when using more elaborated decoders, such as proposed in [20, 21]. For the $0^\circ$ direction, there is a weakly significant improvement from 0 dB to 3 dB ($p = 0.061$) and significant improvements from 0 dB and 3 dB to 6 dB ($p \leq 0.028$). Weakly significant reduction is achieved when increasing the filter gain from 0 dB to 3 dB and from 3 dB to 6 dB ($p \leq 0.096$) for $±45^\circ$, whereas the reduction is significant when increasing the gain from 0 dB to 6 dB ($p = 0.004$). For $±135^\circ$, there is no significant reduction from 0 dB to 3 dB ($p = 0.37$), a weakly significant reduction from 3 dB to 6 dB ($p = 0.08$), but a significant reduction from 0 dB to 6 dB ($p = 0.016$).

Increasing the filter gain from 0 dB to 3 dB yields a weakly significant improvement ($p = 0.052$), whereas the improvement from 0 dB to 6 dB is significant ($p = 0.039$) and the improvement from 3 dB to 6 dB is not ($p = 0.82$).

**Conclusion**

This contribution presented a simple and efficient approach to reduce front-back confusion in binaural rendering. The approach can be applied to any HRIR/BRIR-based rendering without additional computational costs during playback by exchanging the original impulse responses for modified ones. In contrast to the strong level modifications that 360° video players typically apply, our approach exaggerates spectral cues that are already present in the original impulse responses. It can be implemented as a simple shelving filter at 3 kHz that increases high frequencies for frontal directions and decreases them at the rear.

The effectiveness of the approach was evaluated in a listening experiment. For discrete rendering, the amount of front-back confusions could be reduced from 32.5% to 15% (3 dB filter gain) and less than 4% (6 dB filter gain), respectively. Rendering employing 2D first-order Ambisonics improves from 41.25% to 27.5% and 15%. The poorer performance of Ambisonics is caused by the coloration of the simple decoder and can be assumed to improve with more elaborated decoders. Although an additional room simulation using an image-source model improved externalization, it did not reduce front-back confusions.

In general, a high-frequency attenuation of 6 dB can be recommended for rear directions, whereas the boost for frontal directions should not exceed 3 dB to maintain natural timbre.
References


