Simple Reduction of Front-Back Confusion in Static Binaural Rendering

Matthias Frank and Franz Zotter
Institute of Electronic Music and Acoustics, Email: {frank,zotter}@iem.at
University of Music and Performing Arts, 8010 Graz, Austria

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.

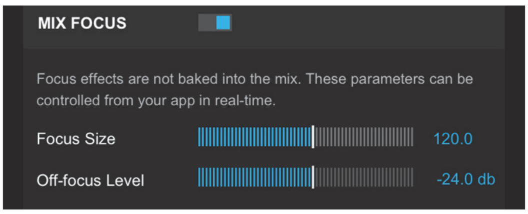

Figure 1: Facebook 360 Spatial Workstation Control Plugin.

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.

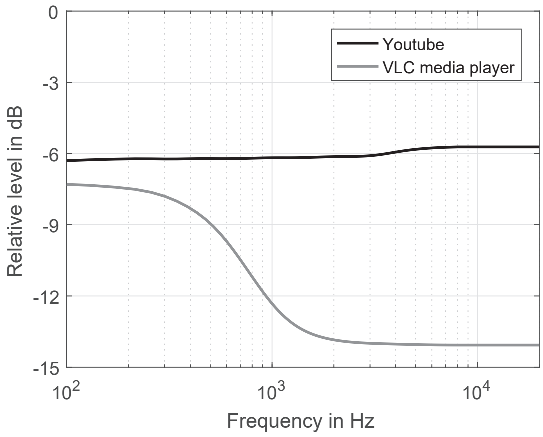

Figure 2: Frequency-dependent attenuation of sound from the back in relation to frontal sound, measured for Youtube and VLC media player.

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.

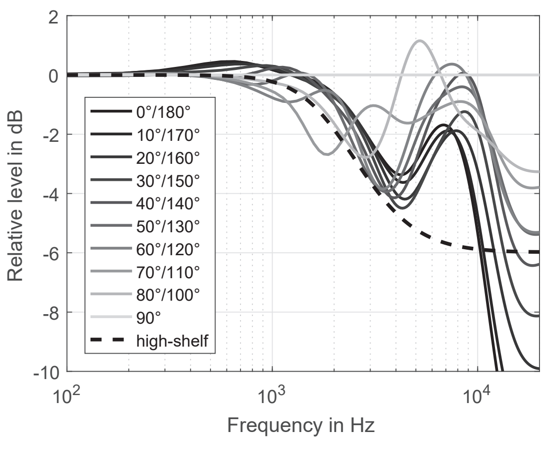

Figure 3: Spectral differences between HRIRs (dataset from [15]) of symmetrical directions in the horizontal plane.

Figure 3 shows that all pairs from 0°/180° to 50°/130° share the similar shape of a high-shelf at 3 kHz. While the peak at 8 kHz increases for more lateral directions, the low-pass effect above 10 kHz becomes weaker. The differences clearly vanish for the pairs closer to the interaural axis. A simple model of the spectral behavior can be a 3 kHz high-shelf filter with a gain of -6 dB for directions behind ±130° that reduces its gain to 0 dB at 90°. This filter avoids additional cues for height in the 8 kHz region [16] and partly keeps the low-pass above 10 kHz. 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 3 dB at the front and attenuated by 3 dB at the back, cf. gray curve for $g_{max} = 3$ dB in Figure 5.

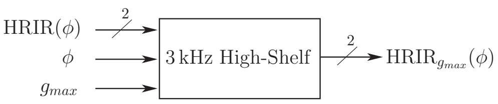

Figure 4: Direction-dependent filter for HRIRs to reduce front-back confusions.

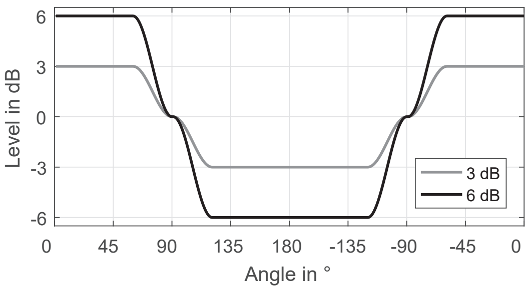

Figure 5: Gain of shelving filter in dependence of angle ϕ for maximum gain g_{max} of 3 dB and 6 dB.

Experiment

The effectiveness of the presented approach is evaluated in a listening experiment using HRIRs from [15] for discrete rendering and 2D first-order Ambisonics (4 virtual loudspeakers at $\phi_L = \{0°, 90°, 180°, -90°\}$), cf. Figure 6. The experiment used a 2-alternative forced choice paradigm and the listeners had to answer whether they perceived the auditory event in front or behind the interaural axis by pressing buttons on a remote control.

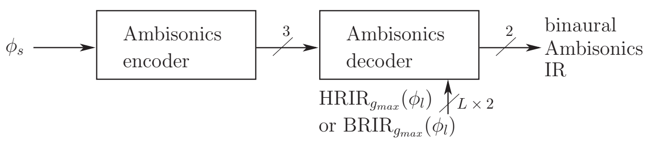

Figure 6: Processing for binaural 2D 1st-order Ambisonics.

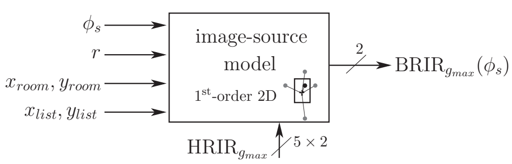

Figure 7: Processing for creating BRIRs from HRIRs.

The German and English recordings from female and male speech of the EBU SQAM database [17] were employed as sound material in the experiment. In order to avoid learning effects/comparison of different rendering conditions and to enforce the internal reference, the sounds were randomly selected. To this end, each of the speech recordings was cut into pieces of 2 to 5 seconds. The 4 directions of $\phi_s = \{0°, \pm45°, \pm135°, 180°\}$ were evaluated for each of the 12 rendering conditions from Table 1. Each condition and direction was repeated 10 times, resulting in 480 decision tasks for each listener. Note that the directions of ±45° and ±135° were each evaluated 5 times for the positive and for the negative direction, respectively.

The whole experiment was divided into 2 parts, one with the dry signals (HRIRs) and one with room information (binaural room impulse responses, BRIRs) from a first-order 2D image-source model, cf. Figure 7. The simulated room had a size of 5 m × 8 m and a frequency-independent reflection factor of 0.5 with the virtual loudspeakers arranged on a radius of 2 m 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.

Renderer	IRs	room	g_{max}
discrete	HRIR	no	0 dB
discrete	$HRIR_3$	no	3 dB
discrete	$HRIR_6$	no	6 dB
discrete	BRIR	yes	0 dB
discrete	$BRIR_3$	yes	3 dB
discrete	$BRIR_6$	yes	6 dB
Ambisonics	HRIR	no	0 dB
Ambisonics	$HRIR_3$	no	3 dB
Ambisonics	$HRIR_6$	no	6 dB
Ambisonics	BRIR	yes	0 dB
Ambisonics	$BRIR_3$	yes	3 dB
Ambisonics	$BRIR_6$	yes	6 dB

Table 1: Binaural rendering conditions in the experiment.

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 6 dB 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.

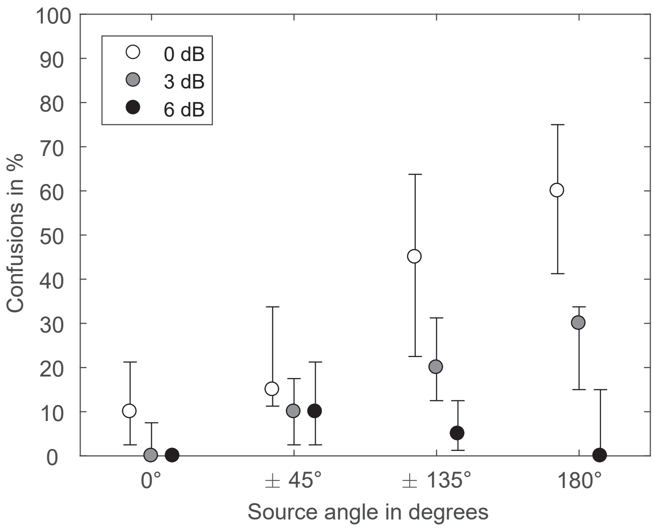

Figure 8: Median values and corresponding 95% confidence intervals of front-back confusions for discrete rendering in dependence of source angle and g_{max} of the shelving filter summarizing HRIRs and BRIRs.

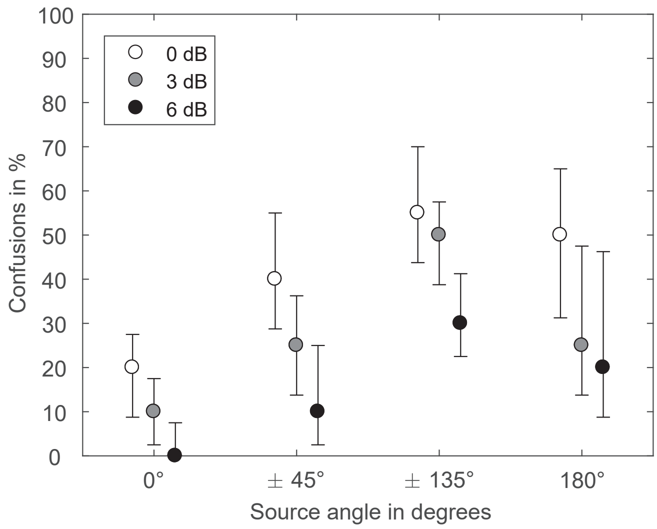

Figure 9: Median values and corresponding 95% confidence intervals of front-back confusions for first-order Ambisonics rendering in dependence of source angle and g_{max} of the shelving filter summarizing HRIRs and BRIRs.

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° direction. For ±45°, 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°, 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° ($p \leq 0.009$).

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 propsed in [20, 21].

For the 0° 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°, whereas the reduction is significant when increasing the gain from 0 dB to 6 dB ($p = 0.004$). For ±135°, 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

[1] J. C. Makous and J. C. Middlebrooks, "Two-dimensional sound localization by human listeners," *The journal of the Acoustical Society of America*, vol. 87, no. 5, pp. 2188–2200, 1990.

[2] R. Bomhardt and J. Fels, "The influence of symmetrical human ears on the front-back confusion," in *Audio Engineering Society Convention 142*, May 2017. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=18590

[3] R. Bomhardt, H. Braren, and J. Fels, "Individualization of head-related transfer functions using principal component analysis and anthropometric dimensions," *Proceedings of Meetings on Acoustics*, vol. 29, no. 1, p. 050007, 2016. [Online]. Available: https://asa.scitation.org/doi/abs/10.1121/2.0000562

[4] K. Sunder, E.-L. Tan, and W.-S. Gan, "On the study of frontal-emitter headphone to improve 3-d audio playback," in *Audio Engineering Society Convention 133*, Oct 2012. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16502

[5] ——, "Individualization of binaural synthesis using frontal projection headphones," *J. Audio Eng. Soc*, vol. 61, no. 12, pp. 989–1000, 2013. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=17076

[6] H. Pomberger, A. Sontacchi, M. Frank, T. Gmeiner, and M. Lucchi, "Improved localization in the median plane with cue-preserving headphones," in *Fortschritte der Akustik, DAGA*, 2018.

[7] F. L. Wightman and D. J. Kistler, "Resolution of front–back ambiguity in spatial hearing by listener and source movement," *The Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2841–2853, 1999.

[8] J. Braasch, J. T. Strong, and N. Xiang, "A binaural model with head motion that resolves front-back confusions for analysis of room impulse responses," in *Audio Engineering Society Convention 129*, Nov 2010. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15701

[9] H. Higuchi, H. Hokari, A. Kudo, and S. Shimada, "An improved method for accurate sound localization," in *Audio Engineering Society Convention 118*, May 2005. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13158

[10] T. Ashby, R. Mason, and T. Brookes, "Head movements in three-dimensional localization," in *Audio Engineering Society Convention 134*, May 2013. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16782

[11] C. Zhang and B. Xie, "Dynamic binaural reproduction of 5.1 channel surround sound with low cost head-tracking module," in *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*, Aug 2014. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=17367

[12] P. Stitt, E. Hendrickx, J.-C. Messonnier, and B. Katz, "The influence of head tracking latency on binaural rendering in simple and complex sound scenes," in *Audio Engineering Society Convention 140*, May 2016. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=18289

[13] M. Romanov, P. Berghold, M. Frank, D. Rudrich, M. Zaunschirm, and F. Zotter, "Implementation and evaluation of a low-cost headtracker for binaural synthesis," in *Audio Engineering Society Convention 142*, May 2017. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=18567

[14] H. Pöntynen, O. Santala, and V. Pulkki, "Conflicting dynamic and spectral directional cues form separate auditory images," in *Audio Engineering Society Convention 140*, May 2016. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=18280

[15] B. Bernschütz, "A spherical far field HRIR/HRTF compilation of the Neumann KU 100," in *Proceedings of the 40th Italian (AIA) Annual Conference on Acoustics and the 39th German Annual Conference on Acoustics (DAGA) Conference on Acoustics*, 2013, p. 29.

[16] J. Blauert, "Sound localization in the median plane," *Acta Acustica united with Acustica*, vol. 22, no. 4, pp. 205–213, 1969.

[17] EBU. (2008) EBU SQAM CD: Sound Quality Assessment Material recordings for subjective tests. [Online]. Available: https://tech.ebu.ch/publications/sqamcd

[18] D. K. Reed and R. C. Maher, "An investigation of early reflection's effect on front-back localization in spatial audio," in *Audio Engineering Society Convention 127*, Oct 2009. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15079

[19] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 94, no. 1, pp. 111–123, 1993.

[20] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural rendering of Ambisonic signals by HRIR time alignment and a diffuseness constraint," *J. Acoust. Soc. Am., submitted*, 2018.

[21] ——, "Perceptual Evaluation of HRTFs with Reduced Spatial Complexity," in *Fortschritte der Akustik, DAGA*, 2018.