Low-frequency trick to improve externalization with non-individual HRIRs

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

Introduction

For headphone rendering of sounds in virtual acoustic environments, we know that externalization as perfect auditory illusion benefits from individual HRIRs [1, 2, 3], a bit from dynamic rendering with head tracking [4, 2, 5], from avoiding the room divergence effect [6, 7], from rendering a natural acoustical room with BRIR lengths of at least 20 or 40 ms [2, 8, 9, 10], spectral detail of the direct sound [11] and ILD fluctuations of natural BRIRs [12]. And yet in many cases, it is impossible to implement all of these steps, but simple tricks prove to be useful [13, 14].

As in a recent binaural downmix produced at IEM[1], low frequencies required boosts up to 20 dB and more to achieve authentic pop musical qualities. There no practical way of involving individual HRIRs in such a production. Moreover, exaggerated low-frequency content might anyway impede the successful illusion of external sounds: Some time ago, a master thesis [15] showed that KEMAR HRIRs, which appear to be high-passed, externalized better than HRIRs from AKG that were more balanced by its stronger low-frequency content; for non-individual HRIRs, there also appears to be a slight tendency of high-passed speech to externalize a bit better in front in [9, Fig.6b].

Attenuation of low frequencies may be acceptable in applications supporting speech communication, but it is unacceptable in audio applications. Neither is the addition of audible reverberation acceptable, whenever it modifies the intended content of the audio production [14]. Summarized requirements: low computational effort, no additional reverberation, no individualization.

In this contribution, we propose to keep but decorrelate frequencies below 250 Hz to improve externalization. By focusing on low frequencies, we aim at keeping the distortion of the room acoustical impression largely unaltered.

Low-frequency effect

The simplistic *Low-Frequency Sanitizer High-Frequency Externalizer* (LFSanHFExt, right in Fig. 1) aims at improving the externalization of high-frequency sounds with non-individual HRIRs without adding reverberation. Still, low-frequency distortions similar to those of natural listening environments need to be accepted.

As mainly the frontal direction is affected from poor externalzation [16, 4], the idea of the approach is to enrich a frontal sound with a sequence of reflected sounds from the left and right.

[1] https://www.youtube.com/watch?v=7XLx-VXwpY0

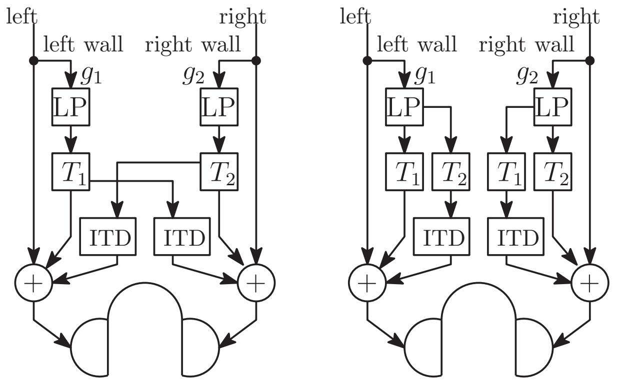


Figure 1: Processing scheme for adding simplified low-frequent left and right wall reflection to a binaural signal (left); the proposed scheme (right) further simplifies cross-talk at low frequencies assuming frontal/dorsal directions.

Fig. 1 sketches the initial idea, where the direct sound is assumed to arrive without delay and attenuation, $g_0 = 1$, from a physical distance r_0. Reflected sounds from the left and right should be arriving at 5% unsymmetrical relative delays of $T_{1,2} = \frac{(r_1-r_0)(1\pm\frac{5\%}{100\%})}{c}$ with the relative gains $g_{1,2} = \frac{r_0}{r_0+(r_1-r_0)(1\pm\frac{5\%}{100\%})}$. Inter-aural time delays are assumed to produce signals corresponding to sounds from the left and right, with ITD=27 samples for both the left and right reflection, implemented at a sample rate of 44.1 kHz. The low-pass filter was a simple first-order low pass at 250 Hz to avoid introducing audible reflections (such would become noticeable with cutoff above 400 Hz). At those frequencies, both ear signals are more or less equal for frontal sounds, so the 2×2 MIMO system (left) is simplified to a SISO system (right) in Fig. 1.

The impulse responses are shown in Fig. 2 for the lengths $r_0 = \{3, 7, 10\}$ m and $r_1 = 17$ m and when $g_{1,2} = 0$ (neutral case). Fig. 3 shows the emerging comb-filter frequency responses for left and right signal and the sum-square response of both. Clearly, the simple implementation does not avoid comb filtering below 125 Hz. There may be more elaborate implementations based on all-pass filters able to avoid such interferences, however, our first results suffered from annoying phasiness so all passes were discarded.

To estimate the effect of the approach, Fig. 4(b) shows the resulting third-octave ILD fluctuations of LFSanHFExt operating on frontal HRIRs, which somehow approximates ILD fluctuations found in binaural room responses Fig. 4(a), below 500 Hz. The listening experiment of the next section investigates which improvement this accomplishes in terms of externalization.

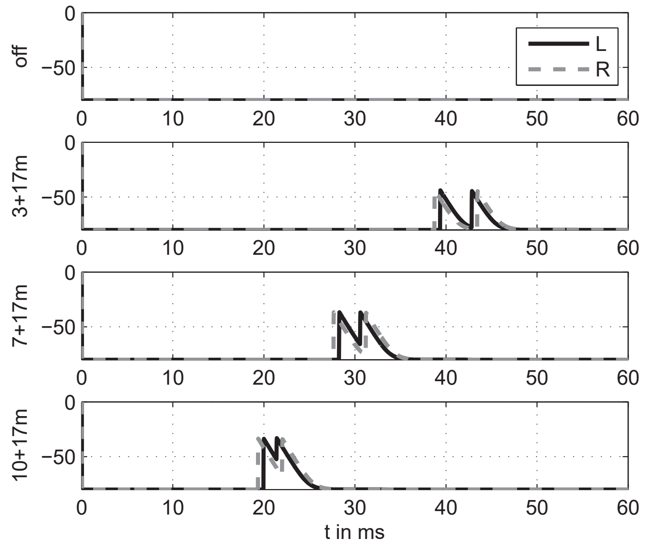


Figure 2: Low-frequency sanitizer/high-frequency externalizer impulse responses in dB with different settings: (i) neutral, (ii) direct 3m plus left/right (17m-3m)±5%, (iii) direct 7m plus left/right (17m-7m)±5%, (iv) direct 10m plus left/right (17m-10m)±5%.

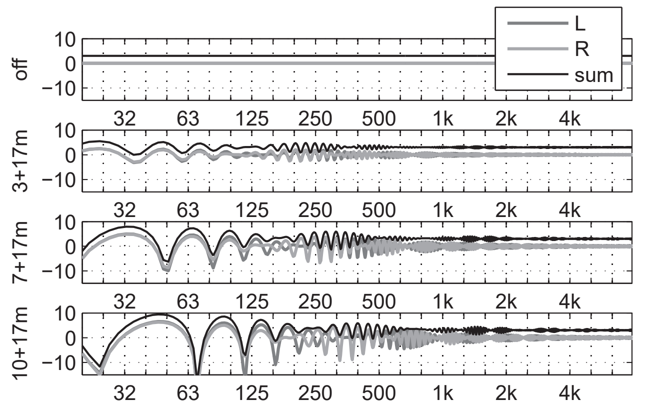


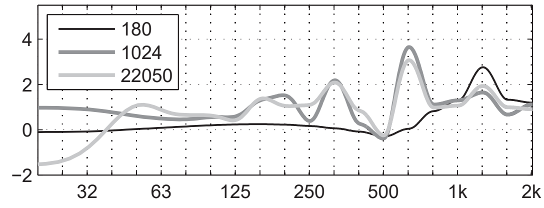
Figure 3: Low-frequency sanitizer/high-frequency externalizer frequency responses (frequency in Hz, magnitude in dB) corresponding to Fig. 2, black: sum-square result, light gray: right, gray: left.

Experiment

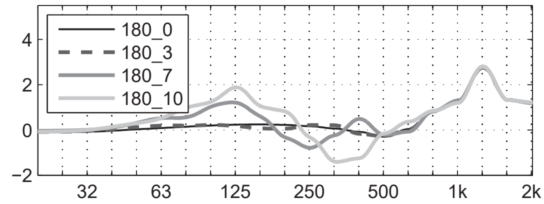
The experiment used the above-presented settings of LFSanHFExt fed by set of binaural signals that already contain the interaural and non-individual spectral cues.

To generate the binaural signals feeding the algorithm, the experiment employed convolution of a mono signal with BRIRs from either the center or right loudspeaker at the IEM production studio measured with a KU100 dummy head. These BRIRs were faded out with a 100-samples $\cos^2$ quarter wave to cut them either to 180, 1024, or 22050 samples to simulate HRIRs and differently long BRIRs.

To avoid the coloration of the headphone used in the experiment, an FIR equalizer was employed using the 2048 sample headphone equalization impulse responses of the AKG K-702 from [17], measured with the KU100 dummy head, see response in Fig. 5.



(a) 3^{rd}-oct. ILDs for different BRIR lengths.



(b) 3^{rd}-oct. ILDs for HRIR under varied effect settings.

Figure 4: ILD for selected BRIR lengths and varied LFSanHFext conditions for the center loudspeaker (frequency in Hz, magnitude in dB).

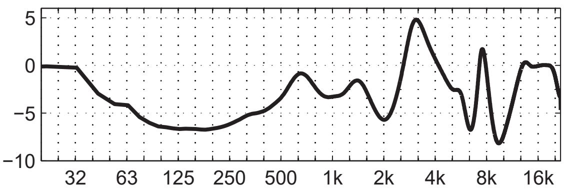


Figure 5: EQ frequency response for AKG K702 open headphones from [17] (frequency in Hz, magnitude in dB).

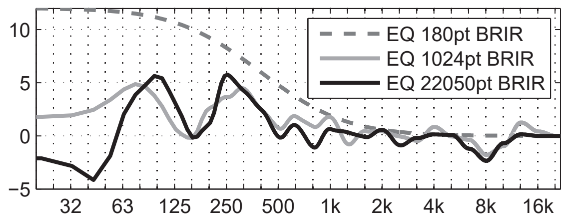


Figure 6: 3^{rd}-octave equalizer with 12 dB bass boost below 200 Hz, and ear-adjusted 3^{rd}-octave noise levels to fit the BRIRs with 1024 and 22050 samples to the 180 sample HRIR.

What is more, the different HRIR/BRIR lengths yield different coloration which was avoided by (i) a simple parametric shelving filter reaching +12 dB below 200Hz for the HRIRs, and (ii) third-octave filters for the differently long BRIRs, as shown in Fig. 6. These filters were aurally obtained by manual equalization of the levels when feeding third-octave noise into the experimental system by both authors. The equalizer was implemented as 4096 sample linear-phase FIR filter bank.

The participants' task was to rate the externalization on a continuous scale with three ticks {*in, on, far from*} *head*. Rating and presentation was done using a graphical user interface for every multi-stimulus task. Each of the multi-stimulus tasks uniformly presented either one of two full-band music signals or a speech signal, uniformly pre-processed either with HRIRs/BRIRs the left or right loudspeaker.

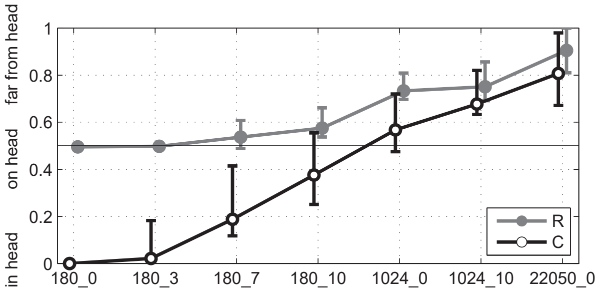


Figure 7: Results of the multi-stimulus listening experiment for different stimuli and the various BRIR/LFSanHFext conditions showing medians and 95% confidence intervals. Black line indicates the *on head* boundary between internalized/externalized.

As the 7 conditions of each multi-stimulus presentation, the signals were processed in 7 different ways:

[180_0] 180 samples HRIR
[180_3] LFSanHFExt $r_1 = 3$ m, 180 samples HRIR
[180_7] LFSanHFExt $r_1 = 7$ m, 180 samples HRIR
[180_10] LFSanHFExt $r_1 = 10$ m, 180 samples HRIR
[1024_0] 1024 samples BRIR
[1024_10] LFSanHFExt $r_1 = 10$ m, 1024 samples BRIR
[22050_0] 22050 samples BRIR

The sequence of (signal)×(L/R)= $3 \times 2 = 6$ multi-stimulus tasks was individually randomly ordered for every participant and contained a randomly arranged set of the 7 differently processed examples. A button allowed to make the examples appear in ascending order according to the current rating to facilitate careful comparison. There were 11 participants of the experiment aged between 23 and 37, and the duration was 15 min on average for everyone to accomplish the 6 multi-stimulus tasks.

Results and Discussion

Fig. 7 shows the analysis of the responses in terms of median values and 95% confidence intervals.

While the frontal presentation using the HRIRs/BRIRs of the center loudspeaker C clearly yields in-head localization for the plain HRIR, the the algorithm in its $r_2 = 10$ m setting clearly moves the localization towards on-head localization, so is able to increase the externalization. The BRIR conditions are superior and it seems that for the frontal direction with the short 1024 sample BRIR, the algorithm further improves externalization.

For the lateral presentation using the HRIRs/BRIRs of the right loudspeaker R, the scale used starts with on-head localization, and the algorithm also provides improvement towards more externalization. Interestingly, it was reported by the participants that the lateral conditions with plain HRIRs alone was not perceived to be at 30° where the right loudspeaker is set up, but at a larger angle about 60 or 70°.

A Kruskal-Wallis test showed that most conditions are different at a high significance level $p < 0.05$. Exceptions are: (i) the 180_0 and 180_3 conditions, (ii) the 180_7 and 180_10 condition for the right HRIR, (iii) the 1024_0 and 1024_10 conditions with a weak effect for the center BRIR.

One can find in literature, e.g. about dry and individualized HRIRs shows [16, Fig. 7] that individualization helps externalization[2], however, frontal and dorsal directions still always appear to externalize less well for dry sounds, similar as in [4]. Despite individual and dry HRIRs achieve externalized localization, the best they achieve often seems to be localization that sticks on the forehead, or does not move far away from it.

Conclusion

We have shown a simple approach to filter non-individual binaural signals and turn in-head localization to at least on-head localization. The suggested algorithm HLSanHFExt produces a low-frequency ILD-manipulation that would also be present in a room. By the restriction of the effect to low frequencies, LFSanHFExt is not audible as additional room reverberation, but it has an influence on localization of low frequencies.

Results suggest that the approach is effective for both HRIRs and short BRIRs. For dry HRIRs, the algorithm appears to modify the externalization towards the one achievable with individual HRIRs.

References

[1] D. R. Begault, E. M. Wenzel, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *Journ. Audio Eng. Soc.*, vol. 49, no. 10, pp. 904–916, 2001.

[2] F. Völk, "Externalization in data-based binaural synthesis," in *Fortschritte der Akustik - NAG/DAGA*, Rotterdam, 2009.

[3] J. Oberem, B. Masiero, and J. Fels, "Experiments on the authenticity and plausibility of binaural reproduction via headphones employing different recording methods," *Applied Acoustics*, vol. 114, pp. 71–78, 2016.

[4] W. O. Brimjoin, A. W. Boyd, and M. A. Akeroyd, "The contribution of head movement to the externalization and internalization of sounds," *PLOS one*, vol. 8, no. 12, 2013.

[5] E. Hendrickx, P. Stitt, J. C. Messonnier, J.-M. Lyzwa, B. F. Katz, and C. de Boishéraud, "The role of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis," *Journ. Acoust. Soc. Am.*, vol. 141, no. 3, pp. 2011–2023, 2017.

[6] S. Werner, G. Götz, and F. Klein, "Influence of head tracking on the externalization of auditory events at divergence between synthesized and listening room using a binaural headphone system," in *prepr. 9690, 142nd AES Conv.*, Berlin.

[2] sets 2/4-indiv. vs. 1/3-non-indiv in the paper

[7] F. Klein, S. Werner, and T. Mayenfels, "Influences of tracking on externalization of binaural synthesis in situations of room divergence," *Journ. Audio Eng. Soc.*, vol. 65, no. 3, pp. 178–187, 2017.

[8] R. Crawford-Emery and H. Lee, "The subjective effect of brir length perceived headphone sound externalisation and tonal colouration," in *prepr. 9044, 136th AES Conv.*, Berlin, 2014.

[9] J. Catic, S. Santurett, and T. Dau, "The role of reverberation.related binaural cues in the externalization of speech," *Journ. Acoust. Soc. Am.*, vol. 138, no. 2, pp. 1154–1167, 2015.

[10] J. Sinker and B. Shirley, "The effect of early impulse response length and visual environment on externalization of binaural virtual sources," in *prepr. 9552, 140th AES Conv.*, Paris, 2016.

[11] H. G. Hassager, F. Gran, and T. Dau, "The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment," *Journ. Acoust. Soc. Am.*, vol. 139, no. 5, pp. 2992–3000, 2016.

[12] J. Catic, S. Santurett, J. Buchholz, F. Gran, and T. Dau, "The effect of interaural-level-difference fluctuations on the externalization of sound," *Journ. Acoust. Soc. Am.*, vol. 134, no. 2, pp. 1232–1241, 2013.

[13] G. Wersényi, "Effect of emulated head-tracking for reducing localization errors in virtual audio simulation," *IEEE TASL*, vol. 17, no. 2, pp. 247–252, 2009.

[14] Y. G. Kim, C. J. Chun, H. K. Kim, Y. J. Lee, D. Y. Jang, and K. Kang, "An integrated approach of 3D sound rendering techniques for sound externalization," in *Pacific-Rim Conference on Multimedia*. Shanghai: Springer, Berlin, Heidelberg, 2010.

[15] M. Noisterning and P. Majdak, "Implementierung von kopfpositionsbezogener Binauraltechnik," M. Thesis, University of Music and Performing Arts, Graz, 2002. [Online]. Available: http://phaidra.kug.ac.at/o:11027

[16] S.-M. Kim and W. Choi, "On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach," *Journ. Acoust. Soc. Am.*, vol. 117, no. 6, pp. 3657–3665, 2005.

[17] B. Bernschütz, "A spherical far field HRIR/HRTF compilation of the Neumann KU 100," in *Fortschritte der Akustik - AIA/DAGA*, Merano, 2013.