The influence of the DRR on audiovisual coherence of a real loudspeaker playing virtually over headphones

Annika Neidhardt[1], Christian.Schneiderwind[1]

[1] *Institute of Media Technology, Technische Universität Ilmenau, 98693 Ilmenau, Germany,*

Email: annika.neidhardt@tu-ilmenau.de, christian.schneiderwind@tu-ilmenau.de

Abstract

One of the challenges in realizing auditory augmented reality is the audiovisual alignment of virtual sound sources created with dynamic binaural synthesis over headphones with a real visual object. In this study the influence of the direct-to-reverberant-energy-ratio (DRR) on the audiovisual coherence is investigated. An AR scenario was realized in which two real loudspeakers are visible and one of them reproduced sound only virtually. A listening test was conducted to evaluate the audible effects caused by a step-wise DRR manipulation of the Binaural room impulse responses (BRIRs). The results indicate that audiovisual plausibility tolerates DRR deviations above JND level. In addition, slightly increased reverberant energy seems to be preferred over the real DRR values.

Auditory Illusions for Augmented Reality

For the realization of Augmented Reality (AR) virtual sound objects are added to the user's real environment. As a consequence the room acoustic properties of the created virtual content has to match the acoustics of the real room at least perceptually. Often, the virtual sound is intended to seemingly originate from a specific real or virtual visible object. Thus, the virtual sound source and the visible object are required to be perceived as audiovisually coherent to create a convincing illusion. Noticeable deviations in the reproduced reverberation can lead to a degradation in perceived plausibility, localization, externalization and/or audiovisual coherence. This may result in a breakdown of the illusion in the worst case.

For AAR systems that are operable in previously unknown environments, an automatic characterization of this environment's acoustical properties is necessary in order to adjust the reproduction accordingly. For an efficient realization it is of interest, how accurately these properties have to be matched to the real environment. This requires psychoacoustic investigations. One essential question is, how selected room acoustical parameters relate to the auditory perception. For example, Klein et al.[1] proposed a method for the perceptual effects and thresholds of deviations in the reverberation time.

The present study focuses on the direct-to-reverberant-energy-ratio (DRR) which is know to be an essential cue for auditory distance perception in rooms. This preliminary study explores the sensitivity of audiovisual plausibility to deviations of the DRR from the perceptually ideal value. For this purpose the original DRR of the measured BRIRs is varied from +6 dB to -9 dB in intervals of 1.5 dB. The manipulations were realized by applying a simple gain factor to the full reverberant tail of the BRIRs. A psychoacoustic experiment is conducted to evaluate the perceptual consequences of the deviations.

BRIR dataset used for binaural synthesis

A seminar room with a size of 9.9 m×4.7 m×3.1 m was chosen for the study. The room has a volume V=144 m^3 and a reverberation time of RT=1.0 s (broad band). BRIRs were measured with a G.R.A.S. Kemar 45ba at nine positions in 25 cm steps along a 2 m-translation line towards a loudspeaker Genelec 1030A. The distance of the various listening positions to the loudspeaker ranged from 1.25 -3.25 m. For each of the nine positions the Kemar was rotated with an electronic turn table to measure BRIRs in 4° steps for the full 360°. Both, the ears for the head-an-torso-simulator and the acoustic center of the loudspeaker were located 1.60 m above the floor. Fig. 1 shows the arrangement of the listening positions. The same procedure was repeated with AKG K1000 headphones opened by 45° placed on the dummy's ears. BRIRs considering the presence of the headphones causing shadowing effects at the listener's ears are of interest in a perceptual comparison with a real loudspeaker, e.g. to evaluate Authenticity [2] or Plausibility [3, 4] of auditory illusions for AR.

In addition, BRIRs were measured for the same constellation of the two loudspeakers and the nine listening positions along the 2m-line in the listening laboratory of the university in Ilmenau (8.7m×7.6m×2.8m, RT = 0.28s). All of the measured BRIRs are part of a larger data set that was published online for scientific purposes [5].

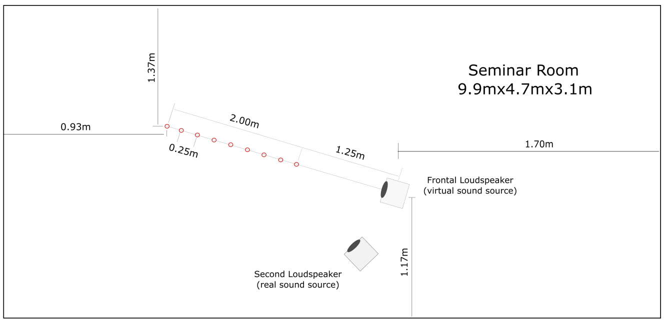

Figure 1: Measurement positions in the seminar room, where the experiment was conducted. The listener can walk along a 2 m-line towards the loudspeaker in the front and turn in arbitrary directions. This speaker reproduced sound only virtually. A second loudspeaker actually reproduced music throughout the experiment.

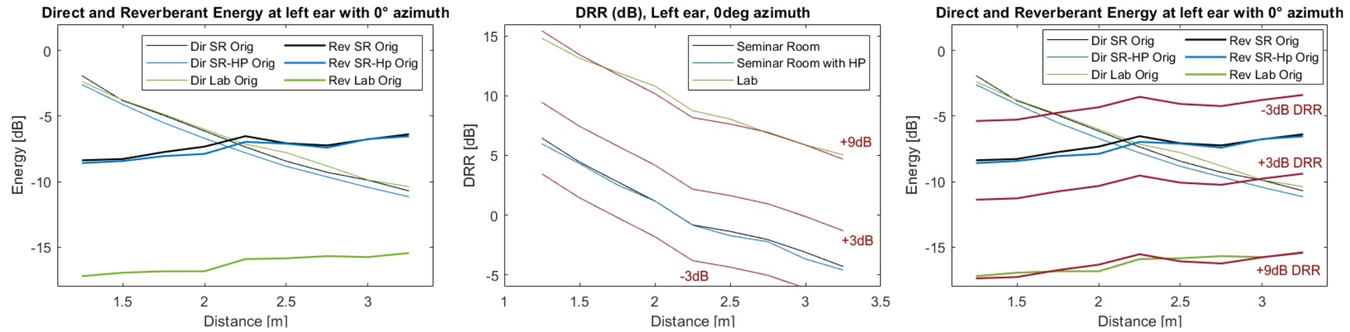

Figure 2: Manipulating the DRR of the measured BRIRs by scaling the reverberation energy - Visualized for BRIRs measured at the left ear for a 0° head orientation (facing the frontal loudspeaker).

Manipulating reverberant energy (DRR)

The DRR is defined by the ratio of direct and reverberant sound energy. Within a room, the DRR varies due to strong distance- and direction-dependent variations of the direct sound energy arriving at the listener. In addition, the reverberant sound energy is likely to vary with position and direction as well. On the one hand, this is caused by the position- and direction-dependent variation of the strong early reflections. On the other hand, the late reverberation is usually not ideally diffuse and may fluctuate [6, 7].
If the same sound source is placed in various rooms and the listener moves in the same relation to the source position, the direct sound arriving at the listener will be the same. Only the room response will change with the room and is likely to vary in its overall energy. Consequently, the DRR will also change with the room due to the different reverberant sound energy.
In this study, the DRR manipulation was realized by an adjustment of the reverberant energy. The first approach consisted of a constant scaling over the entire reverberant part of the BRIRs. For this purpose, the scaling factors were calculated according to the desired deviation in the DRR (in $[dB]$). The second approach was used only for the test case where the BRIRs of the listening lab were scaled to match the DRR of the seminar room BRIRs. Therefore, the temporal progress of energy decay curve (EDC) was adapted using a simple linear scaling curve. The linear curves approximating the EDC of the BRIRs were calculated for both rooms. The ratio of both curves was used as a scaling vector to reshape the reverberant part of the BRIRs of the listening lab.

Psychoacoutic experiment

We conducted an exploratory psychoacoustic experiment to investigate the audible effects caused by the deviations of the reverberant energy and thus of the DRR.

Test setup

The experiment was conducted in the same room and at the same positions, where the BRIRs had been measured, as shown in Fig. 1. The participant had to wear AKG K1000 headphones opened by 45°. An HTC Vive tracker was attached to the headphones to track the motion of the listener's head. The participant was encouraged to walk back and forth on the 2 m-translation line and turn to arbitrary directions. The line was marked on the floor. The frontal loudspeaker used in the measurement was kept at the same position. A second loudspeaker of the same type was placed to its right, with about a 45° angle between both loudspeakers with respect to the frontal position of the translation line. The second loudspeaker reproduced orchestral music without vocals.

Participants

16 people (4 female, 12 male) participated in this study. Their mean age was 33.8 ±3.7 years. According to their statements in the questionnaires, all had normal hearing abilities without any impairments. Half of the test panel may be considered experienced listeners due to their profession as researchers in this field and repeated participation in psychoacoustic evaluations of auditory AR.

Test items and test procedure

20 test cases were considered in this experiment:

- **Seminar room:** -6dB, -4.5dB, -3dB, -1.5dB, Original, +1.5dB, +3dB, +4.5dB, +6dB, +9dB
- **Seminar room with headphones:** -3dB, -1.5dB, Original, +1.5dB, +3dB, +4.5dB, +6dB
- **Listening lab:** Original, -9dB (= DRR of SR), Lab scaled to DRR of SR with time dependent gain

After an active exploration of the scene the subject had to answer the question "Do you have the impression that the loudspeaker you see in the front is playing?" on the rating scale shown in Fig. 3. Each participant had to rate each of the 20 test cases twice in a randomized order. If the participant did not answer with "6 - Clearly Yes" he/she was asked to freely describe which cues limited the impression that the loudspeaker is playing.

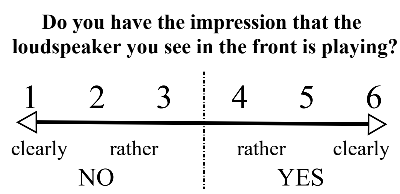

Figure 3: Question and rating scale for the perceptual evaluation in the experiment.

Results

Fig. 4 visualizes the ratings for each test case by the eight experienced listeners (top) and eight inexperienced listeners (bottom). In both plots, the first group of boxes shows the results for the scaled versions of the original measurement in the seminar room (without headphones). A Mann-Whitney-U test was conducted to test for significant differences from the original BRIR set (fifth box). For the experts, in the case of decreased DRR, only at -6 dB the difference was significant. For increased DRR, significant differences can be found starting from a deviation of 4.5 dB. This indicates a quite high tolerance of DRR deviations from the original BRIRs and a slightly higher sensitivity to a decrease of reverberant energy. Comparing all conditions to the -1.5 dB DRR case, which achieved the highest ratings, all other scaled versions were rated significantly lower. Only the ratings of the original BRIRs did not achieve a significance level of 5% ($p=0.057$) in that comparison. This indicates a trend that listeners prefer a slight increase of the reverberant energy. However, the data cannot proof this hypothesis. Furthermore, the results indicate that decreasing the DRR from -1.5 dB to -3 dB leads to significant differences in the audiovisual plausibility. This is interesting because a difference of 1.5 dB is below the JNDs estimated in any previous study. The presence of the second, real sound source could play a role in this context.

The second group of boxes shows the results for the BRIRs measured with the AKG K1000 headphones placed on the dummy's head. Comparing the scaled versions to the original (3rd box in the group), only for the 6 dB DRR case the audiovisual plausibility was significantly degraded. Regarding the original BRIRs without headphones, no obvious advantage can be observed. Significant differences can only be found for the 4.5 dB and 6 dB condition. A pairwise comparison between the various scaled versions with and without headphones shows only a significant difference for the -1.5 dB case. For all of HP-BRIR-scenes the plausibility was rated significantly lower than for the -1.5 dB case without headphones, that achieved the highest ratings.

All test items created from the BRIRs measured in the listening lab did not achieved a comparable plausibility as the original SR scene, although the DRR was scaled to SR values in the second and third lab condition.

For the inexperienced listeners the results are rather undifferentiated and exhibit large confidence intervals. Only extreme conditions produced significant differences. It is questionable whether this test method is suitable for inexperienced/untrained listeners.

Verbal descriptions of perceived differences given by the experienced listeners included *too dry, too close, too small* for conditions with a higher DRR. For 6 and 9 dB DRR change, several participants mentioned *in-head-localization* and the *impression that the sound source moves along* when they walk. Lower DRR conditions were often described as *too reverberant, too distant, too blurry, too wide, too large* in comparison to the visual reference or simply as *behind the loudspeaker.* Although plausibility is significantly degraded for items with a DRR decreased by >3 dB, none of the participants reported affected externalization for these cases.

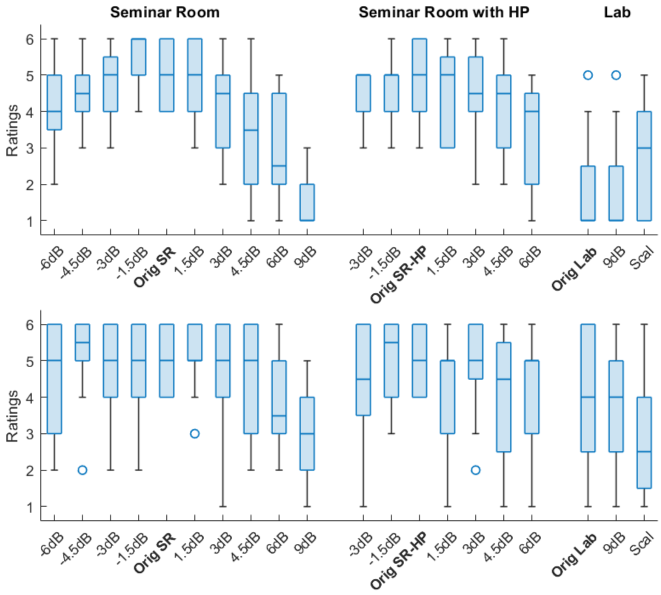

Figure 4: Results for the impression that the loudspeaker is playing as an indirect measure for plausibility rated by experienced (top) and inexperienced (bottom) listeners.

Discussion

The DRR was manipulated by applying a simple, time-invariant gain factor to the reverberant part. This also increases the noise level and decreases the Peak-to-Noise-Ratio (PNR). The BRIRs measured in the seminar room without headphones exhibit a PNR ranging from 60 to 95 dB, depending on position, angle and the side of the ear. For a 0° orientation it ranges between 75 and 85 dB. The DRR manipulation can reduce this PNR by up to 9 dB. According to Hahne et al. [8], in BRIRs for a sound source in the front the noise floor becomes noticeable at around 59 dB. For the other ear the PNR may get as low as about 50 dB for selected angles, but only for one ear at a time. Still, it remains open, whether a listener may notice that at all, since it never occurs at both ears simultaneously. We are confident that this issue did not seriously interfere with the evaluation of audiovisual plausibility.

In this experiment the DRR was varied in discrete intervals. Instead, a test method considering a continuous adjustment by the participant could be of interest to estimate the acceptable range of the DRR in more detail. In an experiment by Werner et al. [9], the participants had the task to adjust the DRR of some BRIRs in a static binaural synthesis to perceptually match it to that of the listening room. For the cases, where the manipulated BRIR was measured in the listening room, listeners are able to adjust the DRR quite accurately. If the BRIR to be manipulated originates from another room with different decay behavior and different early reflection properties, the estimated DRR differed distinctly from that of the listening room.

A similar effect was observed in our experiment for the scenes based on the BRIRs measured in the Lab. Al-

though the DRR of the Lab-BRIRs was adjusted to that of the seminar room (=listening room), the audiovisual plausibility or the impression the visible loudspeaker would reproduce the sound was considerably degraded. Thus, only adjusting the DRR by applying a constant gain-factor in a parametric room auralization will not be sufficient to achieve a perceptual similarity of auralized and actual room. Other properties like the decay behavior and selected properties of the early reflections may not be ignored. This is not new, but essential. Methods like adjusting the (frequency dependent) decay behavior [10, 11] or targeted filtering of a binaural noise sequence [12] have been shown to achieve better room matching results, at least for the late reverberation.
There seems to be a trend that the seminar room scene with -1.5 dB in DRR was perceived more plausible than with the original BRIRs. This hypothesis could not fully be confirmed by the results of the experiment, but it may be interesting to pay more attention to this question in a future study. Furthermore, in our experiment, quite reverberant orchestra music was reproduced by the other loudspeaker to create a kind of real reference. It is also of interest, whether this reverberant signal contributes to such a preference.
Larsen et al. [13] studied the JNDs of DRR in a 2AFC-paradigm, where the participant had to pick the more reverberant sample. The energy of the direct sound was varied to adjust the DRR. In this setup, the estimated JNDs of 2-3 dB for an original DRR of 0 dB and 10 dB and 6-8 dB for 20 and -20 dB. The original DRR values at the investigated listening positions in the seminar room ranged from -4 to +6 dB. Thus, we could expect JNDs between 2-3 dB. In our experiment, audiovisual plausibility was affected starting from -6 dB or +4.5 dB DRR deviation from the original values, which is above the JNDs, as we expected.
Considering the most plausible scene, rather than the original BRIRs, this conclusion does not hold. Decreasing the DRR from -1.5 dB to -3 dB or increasing it from -1.5 to +1.5 dB elicited significant differences of the audiovisual plausibility. This was unexpected as the 1.5 dB difference in DRR is below the JNDs reported so far [13]. A reason could be the visual impression of the scene as well as the real sound source, which both act as a reference in this test paradigm.

Conclusions

Although the presented investigation is only a small preliminary study, quite interesting observations have been made for the influence of DRR deviations on audiovisual plausibility. The test design offers a higher degree of ecological validity for AR scenarios compared to the paradigms used for JND estimation. Manipulating the overall reverberant energy to change the DRR by +4.5 dB or -6 dB affected the audiovisual coherence. Furthermore, slightly increased reverberant energy seems to improve audiovisual plausibility. If BRIRs measured in a different room, a simple adjustment of DRR is not sufficient to achieve audiovisual coherence. Regarding the influence of headphones on the real sound field it was surprising that the participants preferred the BRIRs measured without headphones, despite perceiving the real loudspeaker reproduction through headphones. Inexperienced listeners showed high inconsistencies and partly poor differentiation with this test design. Future work should look into appropriate training approaches or test methods.

Acknowledgement

This study was funded by the Free State of Thuringia, Germany (FKZ: 5575/10-16) and DFG (Project BR 1333/18-1). The authors would like to thank Claudia Stirnat and Tarek Al Sibai for their contributions to the study. We thank all volunteers for their participation.

References

[1] Klein, F., Gari, S.V.A., Arend, J.M., Robinson, P.W.: Towards determining thresholds for room divergence: A pilot study on detection thresholds. IEEE Xplore, Int. Conf. Immersive & 3D Audio, Online, Bologna Italy, 2021.

[2] Brinkmann, F., Lindau, A., Weinzierl, S.: On the authenticity of individual dynamic binaural synthesis. J. Acoust. Soc. Am. 2017 Oct; 142(4):1784-95, doi: 10.1121/1.5005606

[3] Lindau, A., Weinzierl, S.: Assessing the plausibility of virtual acoustic environments. Acta Acustica United with Acustica 2012 98(5):804-810, doi: 10.3813/AAA.918562

[4] Neidhardt, A., Zerlik, A.M.: The availability of a real hidden reference affects the plausibility of position-dynamic auditory AR. Front. VR 2021. doi: 10.3389/frvir.2021.678875

[5] Neidhardt, A., Zerlik, M., Kamandi, S.: BRIR data set for interactive listener translation in two rooms (1.0), Data set (2020). Zenodo. doi: 10.5281/zenodo.3457782

[6] Mijić, M., Masovic, D.: Reverberation Radius in Real Rooms. Telfor Journal, Vol 2, No. 2, 2010.

[7] Alary B., Massé, P., Schlecht, S., Noisternig, M., Välimäki, V.: Perceptual analysis of directional late reverberation. J. Acoust. Soc. Am. 2017 Oct; 149(5):1784-95, doi: 10.1121/10.0004770

[8] Hahne, W., Erbes, V., Spors, S.: On the Perceptually Acceptable Noise Level in Binaural Room Impulse Responses. 45th DAGA Conf., Rostock, Germany 2019.

[9] Werner S, Klein F, Sporer T.: Adjustment of the Direct-to-Reverberant-Energy-Ratio to reach externalization within a binaural synthesis system. Int. AES Conf. on Audio for VR & AR, Los Angeles, CA, USA, 2016.

[10] Jot, J.M., Lee K.S.: Augmented reality headphone environment rendering. AES Int Conf Audio for Virt & Augm Reality, Los Angeles, CA, USA, Sept 2016.

[11] Melchior, F., De Vries, D.: On the visualization and modification of room impulse responses for sound design. 34th Annual German Congress on Acoustics (DAGA), Dresden, Germany 2008.

[12] J.M. Arend, S.V.A. Garí, C. Schissler, F. Klein, P. W. Robinson: Six-Degrees-of-Freedom Parametric Spatial Audio Based on One Monaural Room Impulse Response, J. Audio Eng. Soc., 2021, 69(7/8), 557-575, doi: 10.17743/jaes.2021.0009

[13] Larsen, E., Iyer, N., Lansing, C.R. and Feng, A.S.: On the minimum audible difference in direct-to-reverberant energy ratio. J Acoust Soc Am 2008 Jul; 124(1):450-461.