vdt Verband Deutscher Tonmeister e.V.

AUDIO AES

AES Reviewed Paper at Tonmeistertagung 2018

Presented * by VDT in cooperation with the
Central European Region of the Audio Engineering Society (AES).

Investigations on the Impact of Distance Cues in Virtual Acoustic Environments

Christoph Pörschmann[1], Johannes M. Arend[1,2], Philipp Stade[1,2]
[1] *TH Köln, Institute of Communications Engineering, Cologne, Germany*
[2] *TU Berlin, Audio Communication Group, Berlin, Germany*

Abstract

In this paper, we analyze different auditory distance cues in dynamic binaural synthesis. We compare the contributions of sound intensity, direct-to-reverberant ratio (DRR), and near-field cues. For the auralization, we use the BinRIR method, which allows to generate binaural room impulse responses (BRIRs) for dynamic binaural synthesis based on one measured omnidirectional room impulse response (RIR). With BinRIR, applying a simple geometric model, the listener position can be freely adjusted and the distance cues can be adapted separately. Furthermore, near-field head-related impulse responses (HRIRs) can be applied for direct sound and early reflections if the listener is very close to the virtual sound source. In a listening experiment, we presented stimuli at different distances in four synthesized rooms. In one condition, the stimuli contained natural distance-dependent intensity cues, and in another condition, the stimuli were normalized in loudness. The results showed that even for loudness-normalized stimuli, an adequate distance perception can be obtained by adapting the DRR. The influence of near-field HRIRs, which were also tested in the experiment, is weak.

1. Introduction

Sound intensity, direct-to-reverberant ratio (DRR), as well as the spectrum influence auditory distance perception of sound sources [1] [2] [3]. Furthermore, it is contrarily discussed in literature if and to what extent binaural cues are relevant for the perceived sound source distance in the near field [4] [5]. Generally speaking, perceived source distance increases with decreasing level. In anechoic conditions, the relationship between level and distance from a sound source to the receiver is characterized by the 6 dB law for each doubling of the source distance. In reverberant conditions, this decrease is reduced due to reflections and reverberation. However, as intensity also depends on the source signal, it has to be regarded as a relative distance cue. Thus, the presented signal needs to be compared to a reference in order to judge distance.

In reverberant conditions, the ratio of energy reaching a listener on the direct path to the energy reaching the listener via the reflecting surfaces (DRR) is inversely related to the distance of the sound source. As the DRR is independent of the source signal, it provides absolute distance information [2]. Furthermore, changes of the DRR are typically related to a modified initial time delay gap (ITDG), which describes the temporal difference between the direct sound and the first strong reflection.

Distance perception is generally most accurate when both DRR and level cues are available. However, when analyzing the cues isolated, intensity cues provide more accurate information than DRR only [2]. In highly reverberant environments, though, both cues can provide equally accurate information for distance discrimination. On the contrary,

[*] Please note that the AES Reviewed Papers at Tonmestertagung can be published by both, AES and VDT, in print, online and as PDF download.

ISBN 978-3-9812830-9-9

studies analyzing just-noticeable-differences (JNDs) showed that the JNDs for the DRR are comparably large. According to [2] they are in a range of 2–3 dB for 0 and +10 dB DRR and about 6–8 dB at -10 and +20 dB DRR. This suggests that the principal role of the DRR is to provide a rough absolute distance information, rather than to support fine distance discriminations which can be detected from small changes in overall intensity [3]. According to the the pressure-discrimination hypothesis, the JNDs in source distance are determined by the ability to discriminate changes in sound pressure. According to [6] for broadband noise, the smallest detectable change in level is approximately 0.4 dB.

The spectrum of the signal serves as another distance cue. For larger distances (sound source distances more than 15 m) dissipation causes a low-pass filtering characteristic which increases towards higher distances and by this influences perceived distance. Thus sounds with decreased high-frequency components relative to low-frequency components are perceived to be further away [2]. As this cue is strongly influenced by the spectrum of the source signal, it can be regarded as a relative distance cue. In the near field (sound source distance less than 1 m) the spectrum also changes with distance because diffraction and head-shadowing effects lead to a low-pass filtering characteristic which increases towards lower distances [7] [8]. Additionally, for close sound sources diffraction and head shadowing induce a distance-related change in interaural time differences (ITDs) and interaural level differences (ILDs) [7] [8]. While the influence of distance on the ITDs is relatively low, ILDs change substantially over distance.

Distance estimation is highly relevant for headphone-based virtual acoustic environments (VAEs) which are applied in various areas like audio engineering, telecommunications, or architectural acoustics to create a natural room impression. Furthermore, the use of VAEs is well suited for applying listening experiments and presenting stimuli from different distances and rooms in a natural environment. For example, in the context of distance estimation Zahorik [9] analyzed contributions of different cues to distance estimations based on one measured room. Brungart [10] assessed distance estimation of near-field sound sources in a VAE.

Varying sound source distances can be considered relatively easy in simulation-based systems (e.g. [11]). However, depending on the application, the auralization often relies on binaural room impulse responses (BRIRs), which are either measured with artificial heads for different head orientations or rely on an arbitrary BRIR synthesis. If the BRIRs are acquired with an artificial head, measurement times for the acquisition of a circular (or even a spherical) set of BRIRs are often quite high and increase even more if data at different listening positions and distances needs to be captured. Thus, the use of parametric models allowing to control DRR, sound intensity, and near-field cues separately in order to consider distance modifications on measured data can be beneficial. In a previous publication, we presented such an approach called BinRIR [12] [13]. This algorithm allows to generate BRIRs for dynamic auralization based on one omnidirectional room impulse response (RIR). The approach uses a parametric model for the BRIR synthesis and spatializes an omnidirectional RIR by extracting direct sound, early reflections, and diffuse reverberant parts from the RIR. The synthesis allows for a free shift of the listener position in the room. Additionally, near-field head-related impulse responses (HRIRs) can be applied if the sound source or one of the reflections are very close to the listener. Thus, this approach can be used to auralize arbitrarily chosen listener positions in the virtual room. This allows to create naturally sounding stimuli from measured data and to separately control the different cues which affect distance perception.

Applying BinRIR, we investigated the impact of intensity, DRR, and near-field cues on distance estimation in a listening experiment. In the form of a multi-stimulus comparision test, we tested distance perception for four rooms and different distances between source and listener. Additionally we varied the presented distance cues. The results of the experiment show that intensity and DRR have a significant impact on distance estimation. Even for loudness-normalized stimuli, listeners were able to judge distance appropriately. Near-field cues, however, did not significantly support distance estimation.

The paper is organized as follows: Section 2 describes the BinRIR algorithm and explains how the listener position shifts and the near-field HRIRs are applied. This algorithm provides the basis for the psychoacoustic study. Section 3 presents the listening experiment in which we examined distance estimation for four different rooms. In section 4, we analyze in which way the different cues contribute to distance estimation. Finally, section 5 concludes the paper.

2. BinRIR Method

For the listening experiment presented in this paper, we synthesized stimuli with the BinRIR method [12] [13]. BinRIR generates BRIRs based on a measured omnidirectional RIR and allows to separately modify different distance cues. The basic structure of BinRIR is shown in Fig. 1. The method only applies to frequency components above 200 Hz. For lower frequencies, the interaural coherence of a typical BRIR is nearly one and the omnidirectional RIR can be maintained.

BinRIR requires only one measured omnidirectional RIR to synthesize an arbitrary BRIR dataset. To obtain a BRIR, we use predictable information from geometrical acoustics as well a perception-motivated simplified description of the diffuse sound field. For this, the RIR is split into two different parts. Onset detection is used to identify the direct sound in the ommnidirectional RIR. This section starting with the onset is windowed (5 ms followed by 5 ms raised cosine offset ramp). The following time section is assigned to the early reflections and the transition towards the diffuse reverberation. 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 of the sliding window are determined and assigned to geometric reflections. Each

* Please note that the AES Reviewed Papers at Tonmestertagung can be published by both, AES and VDT, in print, online and as PDF download.

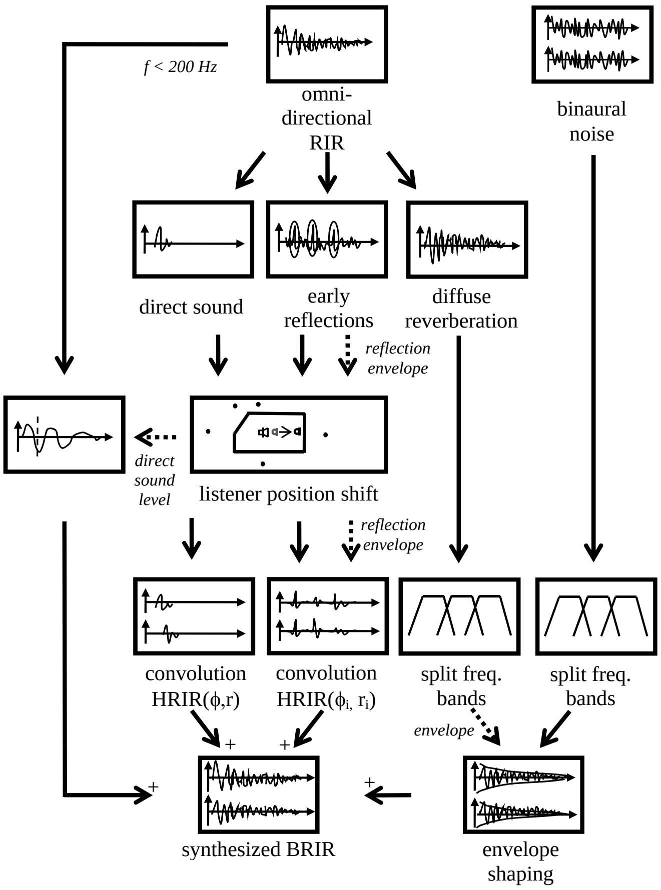

Fig. 1: Block diagram of the BinRIR algorithm for synthesizing a BRIR based on a single omnidirectional RIR.

section comprises both the amplitude and the spectral shape of the reflection. Thus, all specific characteristics of the reflections, e.g. edge diffraction and the acoustical properties of the surfaces, are adequately considered. In contrast, the incidence directions of the synthesized reflections are chosen by the algorithm and base on a spatial reflection pattern adapted from a shoebox room with non-symmetric positioned source and receiver. Each windowed section of the RIR is convolved with the appropriate HRIR. For this, a measured set of HRIRs of a Neumann KU100 artificial head [14] is used. To obtain interim directions between the given directions of the HRIR set, an interpolation by means of spherical harmonics transform is performed [15]. By this a binauralized version of the early geometric reflections is obtained.

All sections of the RIR which were not detected as geometric reflections are assigned as diffuse. Several studies have shown that after the so-called perceptual mixing time, the measured (binaural) room impulse response can be simplified [16][17]. In BinRIR the diffuse reverberation part is synthesized by shaping binaural noise according to the envelope of the diffuse part of the omnidirectional RIR. The diffuse part is considered reaching the listener temporarily and spatially equally distributed and is not influenced by shifts of the listener position. Following recent studies (e.g. [18] [19]) BinRIR overlaps sections with early reflections and diffuse parts instead of using a fixed mixing time. Cross-fading is applied in order to merge sections with geometric reflections and sections with diffuse reverberation.

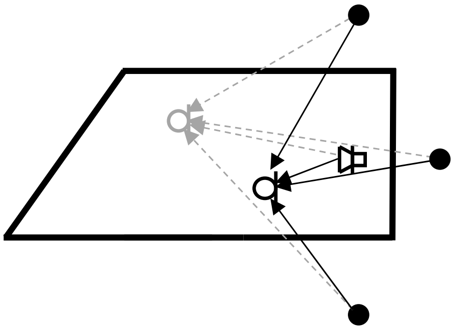

Figure 1: Basic principle of the listener position shifts (LPS) applying mirror images: The amplitude as well as the temporal structure of the direct sound and the early reflections are adapted accordingly. The receiver is moved from an initial position (grey) to a modified position (black). By this the paths of direct sound and reflections are changed.

Fig. 2: Basic principle of the listener position shifts: Mirror images are applied demonstrating the modification of the amplitude and the temporal structure of direct sound and early reflections. The receiver is moved from an initial position (grey) to a modified position (black). By this, the paths of direct sound and reflections are changed.

The possibility to consider listener position shifts in BinRIR based on a simple geometric model is of specific relevance for this study. For this, the distance between the listener and each of the mirror images is determined based on the delay between the corresponding peak of the reflection and the peak of the direct sound. 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 or the reflections (Fig. 2). Thus by shifting the listener position the amplitudes and time delays of the direct sound and the reflections are affected. As a consequence the ITDG and the DRR are modified as well by the listener position shifts.

Furthermore, HRIRs measured at different distances can be considered in BinRIR. Thus, in addition to the used far-field HRIRs [14], near-field HRIRs, as well measured with a Neumann KU100 at distances between 0.25 m and 1.50 m are used [8]. Regarding the distance of the direct sound and each of the reflections, the HRIR is chosen from the set that fits the distance of the reflections best. However, for distances above 2 m, far-field HRIRs are chosen.

The BRIR synthesis is repeated for constant shifts in the azimuth angle (e.g. 1°) for the direct and the reflected sound. Thus, a circular set of BRIRs is created, which can be used for dynamic binaural synthesis. Since binaural cues are absent in the measured omnidirectional RIR, the BinRIR algorithm incorporates several inaccuracies and errors. For example, the directions of incidence of the synthesized early reflections are not in line with the original ones. Hence, differences in perceptual spatial properties (e.g. envelopment) between the original and the synthesized room may occur. In [12] [13] we already performed a detailed technical analysis as well as a perceptual comparison to a measured reference and examined the performance of the algorithm. Thus these issues are not further discussed and analyzed in the present paper.

[*] Please note that the AES Reviewed Papers at Tonmestertagung can be published by both, AES and VDT, in print, online and as PDF download.

3. Listening Experiment

We performed a multi-stimulus comparison test. In two different sessions the listeners estimated the auditory distances of normalized and non-normalized stimuli. By this, we examined the impact of source intensity, DRR, and near-field cues to distance estimation. Generally, intensity cues and DRR both significantly contribute to distance estimation [2] [3]. However, the influence of DRR and intensity strongly depend on the room and thus, it was an essential part of the study to compare the contributions for acoustically different rooms. Finally, the listening experiment served to test if the synthesis of the BinRIR algorithm allows to adapt the perceived distance appropriately. Thus, we included stimuli synthesized with a measured BRIR in the test, for which we already found a good perceptual correlation to the synthesis with BinRIR [20].

3.1. Participants

In total, 34 adults (5 female, 29 male) aged between 18 and 46 years (M = 24.6 years, SD = 5.07) took part in the experiment. Most of them were students at TH Köln. All listeners already participated in previous listening experiments and thus were familiar with dynamic binaural synthesis. None of the subjects reported hearing problems.

3.2. Setup

The experiments took place in the anechoic chamber at TH Köln, which ensured a low background noise level of less than 20 dB(A). Furthermore, this room is completely different to any of the investigated rooms and potential influences of audiovisual room convergence were thus avoided. The experiment was set up, controlled, and executed running the software Scale [21]. The listeners had to enter their ratings on a touch-screen computer (iPad). The stimuli were presented with dynamic binaural synthesis using the SoundScape Renderer [22] and a Polhemus Fastrak head tracking system to consider full circle horizontal head movements with a resolution of 1°. The stimuli were presented via AKG K-601 headphones. No headphone equalization was applied to the stimuli.

3.3. Materials

Rooms We used measurement data from four different rooms [23]. Table 1 gives an overview about the rooms and their properties. The Control Room 7 (CR7) is the main control room for radio drama production at WDR Broadcast Studios in Cologne. We measured the RIRs in the sweet spot, right in front of the mixing console. The Large Broadcast Studio (LBS) and the Small Broadcast Studio (SBS) are as well located at the WDR in Cologne and are used for various recordings of concerts and performances. TGC is a training room of a local dance club in Cologne. The loudspeaker types varied between the four rooms. In CR7, we used the installed Bowers & Wilkens 803D loudspeakers. In LBS and SBS, the source was a full PA stack involving an AD Systems Stium Mid/High unit combined with three AD Systems Flex 15b subwoofers. In TGC, one AD Systems Flex 15 speaker was applied. Please refer to [23] for further information on the geometry of the rooms and the positions of the loudspeakers and the listener. The measurements comprised both, a circular reference set of BRIRs and an omnidirectional impulse response, measured at the pivot position of the artificial head. In this study, the binaural measurements, which were performed in steps of 1° on the horizontal plane with a Neumann KU100 artificial head, were only used to compare the results of the BinRIR synthesis to a binaurally measured reference. A Microtech Gefell M296S microphone was used for the omnidirectional measurement in TGC; for the other omnidirectional measurements, an Earthworks M30 microphone was applied.

Distances BRIR sets describing different distances between sound source and receiver were synthesized with the BinRIR algorithm. The listener position was shifted as described in section 2 according to so-called distance factors (DFs) on a line between sound source and listener. The DFs correspond to the quotient between the synthesized stimulus distance and the distance of the measured RIR. Thus, DF = 1 is equivalent to the measuring distance between source and receiver $\text{Distance}_{\text{SrcRec}}$ of the measurement (see Table 1). For each room, we tested the following DFs: 0.125, 0.25, 0.5, 1, and 2. Additionally, the measured BRIR – in the following named original – was included in the experiment. Moreover, for close distances in which near-field cues might influence distance estimation, a variant assessing the appropriate near-field HRIR for the direct sound and another variant using only far-field HRIRs were synthesized. Depending on $\text{Distance}_{\text{SrcRec}}$, a varying number of near-field stimuli was used for the different rooms. For LBS, we presented only one stimulus using near-field HRIRs, for SBS and TGC two, and for CR7 three near-field stimuli.

Direct-to-reverberant ratio (DRR) The DF strongly influences the DRR. We calculated the DRR as the ratio of the energy reaching the listener within the first 7.5 ms (corresponding to a distance of 2.55 m) after the direct sound

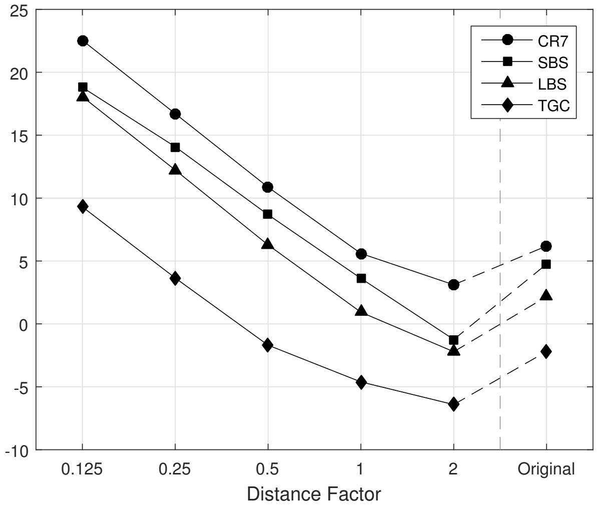

Fig. 3: Direct-to-reverberant ratios (DRRs) for the four different rooms and the varying distance factors (DFs). The highest DRRs can be observed for the rooms with the shortest RT_{60}. In CR7, all synthesized positions are inside the critical distance. On the contrary, for TGC, only the DF = 0.125 and DF = 0.25 are inside the critical distance.

* Please note that the AES Reviewed Papers at Tonmestertagung can be published by both, AES and VDT, in print, online and as PDF download.

Room	Volume	Area	RT$_{60}$	Distance$_{\text{SrcRec}}$
Control Room 7 (CR7)	168 m^3	60 m^2	< 0.25 s	2.7 m
Large Broadcast Studio (LBS)	6100 m^3	579 m^2	1.8 s	13.0 m
Small Broadcast Studio (SBS)	1247 m^3	220 m^2	0.9 s	7.0 m
TGC Training Room (TGC)	191 m^3	67 m^3	2.3 s	6.8 m

Tab. 1: Main properties of the measured and synthesized rooms. The RT$_{60}$ corresponds to the mean reverberation time in the frequency range between 500 and 1000 Hz.

and the energy of the reverberant parts. As only the time differences between the direct sound and the floor reflection is smaller than this value, all the geometric reflections from walls and ceiling are in this context considered as a part of the reverberation. The results for the different rooms are illustrated in Fig. 3. It can be observed that the DRR shows maximal values of up to 22.5 dB for CR7 and generally increases towards small distances. Thus, it can be assumed that in the CR7, especially for nearby sound sources, the influence of the room on distance perception is small. On the contrary, in TGC the DRR is in a range from -6.4 dB to 8.7 dB for all conditions, which probably allows the listener to exploit DRR for distance estimation in all conditions. The influence of near-field cues on the DRR are not shown in Fig. 3. However, they did not exceed 1.1 dB.

Stimuli A looped drum and a guitar sequence were used as test signals. These signals have already been used in previous listening experiments [e.g. 24]. In the listening experiment we presented two conditions of each stimulus. In one condition, the stimuli were loudness normalized according to ITU-R BS. 1770 [25], the other one comprised loudness differences with increasing sound pressure levels towards lower distances as calculated by the geometric model. The playback-level was calibrated to a Leq of 60 dB(A) SPL for the normalized stimuli and for the stimuli with a DF of 1. A maximal Leq of 75.2 dB(A) SPL was obtained for a DF of 0.125 in CR7. The Leq(A) for non-normalized stimuli depending on room and DF is given in Fig. 4. The levels vary most for CR7, especially towards lower distances. In total, level differences of more than 15 dB can be observed here. In TGC, which was the most reverberant room in our study, the level differences between all distances were less than 4 dB. The test signal had only small influence on the SPL; differences between drums and guitar were mostly below 1 dB. Only for low distances in CR7, some differences are about 2 dB. The use of near-field HRIRs did as well only marginally influence the SPL. Maximal deviations of less than 1 dB were observed here.

3.4. Procedure

We performed a multi-stimulus comparison test. In two different sessions, the participants estimated the distances of the normalized and the non-normalized stimuli on a seven-point category scale with the categories 'very close', 'close', 'rather close', 'medium', 'rather distant', 'distant', 'very distant'. The subjects were allowed to rate interim values between the given categories. As in earlier experiments [20] the equidistance between the categories was underlined by the visual presentation. One half of the subjects started with the conditions with the normalized stimuli, the other half with the non-normalized stimuli. In each session, every participant had to rate the stimuli for the four different rooms and the two different test signals. The order of conditions was randomized within each session. For each condition we presented the stimuli for the different DFs, the near-field stimuli, and the binaurally measured reference in one multi-stimulus comparison. The scale was displayed on the tablet computer (iPad) and results were given by setting a slider to the appropriate position. Several test sliders were shown at the same time and the subjects were able to switch between the corresponding stimuli of the DFs as often as required. No additional information on the type, shape or size of the rooms or on the reference was given to the subjects. The experiment started with a short introduction including several test trials in order to make the subjects familiar with the test procedure and the stimuli. It included stimuli with the largest and smallest simulated distance and was used for anchoring as well.

4. Results

We investigated the impact of sound intensity, DRR, and near-field cues on distance estimation. Generally, the subjects stated that distance estimation of the stimuli was easy to perform. However, some participants reported that a few of the close stimuli were perceived being unnatural regarding their tone color or their naturalness.

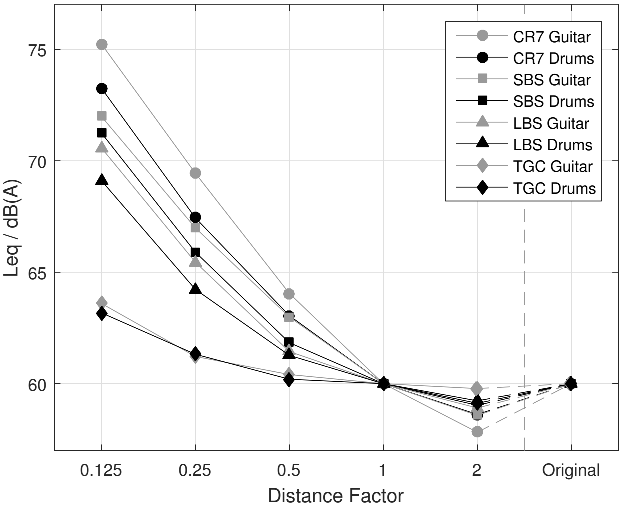

Fig. 4: Levels (Leq) in dB(A) of the non-normalized stimuli depending on the distance factor (DF) for the four rooms. The Leqs vary most for the CR7, especially towards smaller distances, and are minimal for the TGC room at higher distances. Differences between the guitar (grey) and drums (black) source signal do not exceed 2 dB.

* Please note that the AES Reviewed Papers at Tonmestertagung can be published by both, AES and VDT, in print, online and as PDF download.

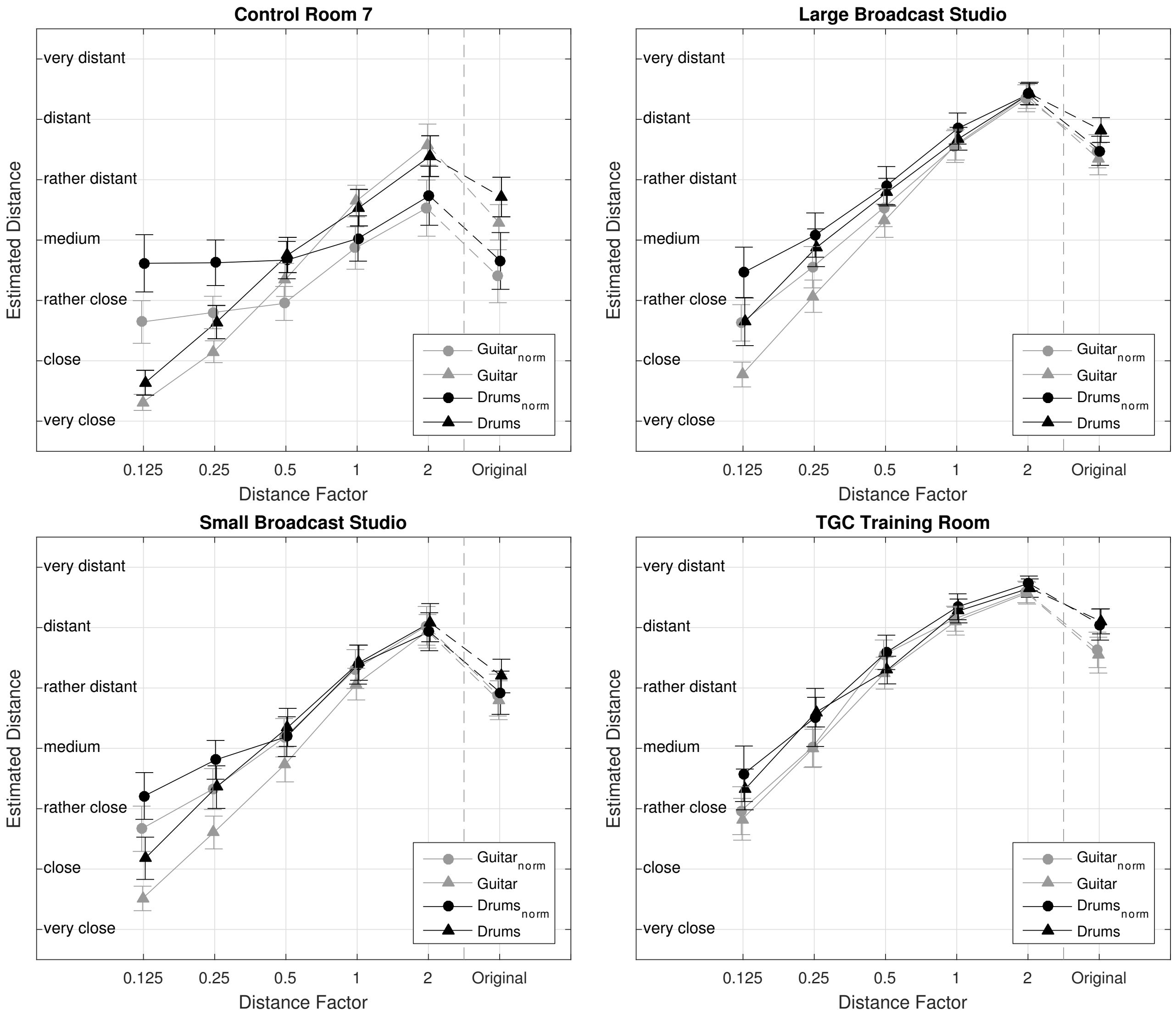

Fig. 5: Estimated distance (mean values and 95 % confidence intervals) for Control Room 7, Large Broadcast Studio, Small Broadcast Studio, TGC Training Room Cologne. On the x-axis, the different distance factors (DF) and the original measurement are plotted.

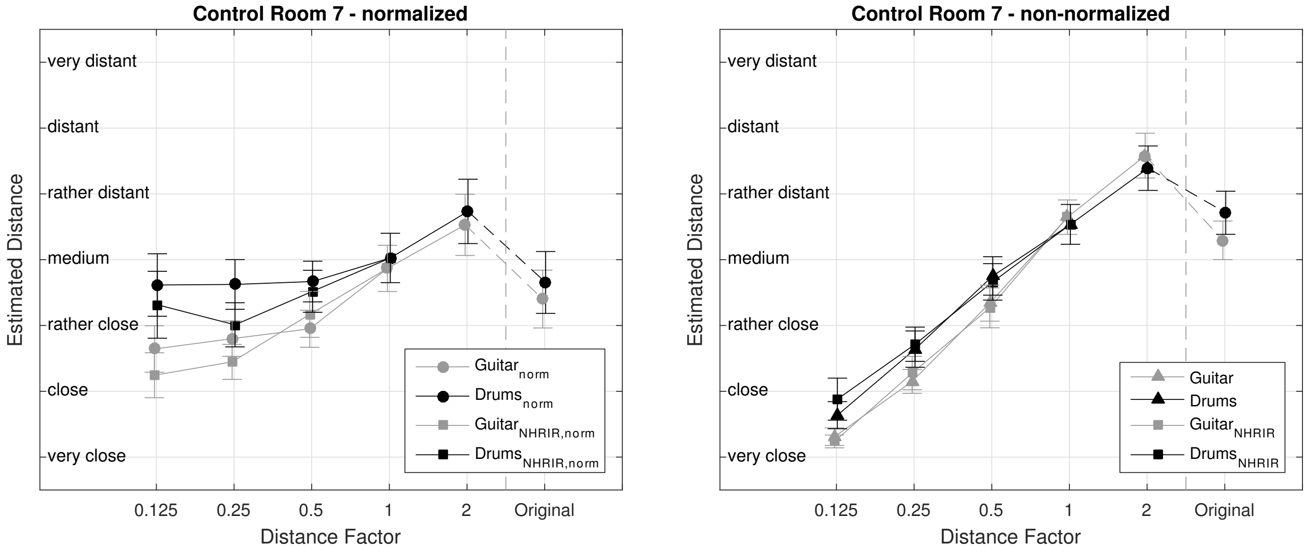

Fig. 6: Impact of near-field HRIRs on estimated distance (mean values and 95 % confidence intervals) for Control Room 7. On the x-axis, the different distance factors (DF) and the original measurement are shown. The left plot shows the loudness-normalized condition, the right one the non-normalized conditions with natural intensity differences.

[*] Please note that the AES Reviewed Papers at Tonmestertagung can be published by both, AES and VDT, in print, online and as PDF download.

First, we performed a Greenhouse-Geiser corrected [26] repeated measures ANOVA with the factors DF, room, test signal, and loudness normalization. A varying number of stimuli assessing the near-field HRIRs was presented for the different rooms. Thus, in order to get a balanced test design, these stimuli were not part of the ANOVA. The analysis yielded significant main effects for all factors. Main factors with highest effect size were DF ($F(4, 132) = 448$, $p < .001$, $\eta_p^2 = .93$) and room ($F(3, 99) = 137$, $p < .001$, $\eta_p^2 = .81$). The effect sizes of test signal ($F(1, 33) = 36.1$, $p < .001$, $\eta_p^2 = .52$) and normalization ($F(1, 33) = 25.5$, $p = .003$, $\eta_p^2 = .44$) were lower. Surprisingly, the main effect of normalization had the lowest effect size. Furthermore, various significant interaction effects were observed. Strongest was the interaction between DF and normalization ($F(4, 132) = 59.0$, $p < .001$, $\eta_p^2 = .64$). Thus in a next step, we further analyzed the data by performing a nested repeated measures ANOVA for each room. Here, we found no significant effect of normalization for TGC ($p = .13$). For SBS ($p < .001$) and LBS (p=.008), there was a significant main effect of normalization which, however disappeared after a subsequent removal of the closest distances from the analysis. We performed the ANOVA for LBS without the DF of 0.125 and for SBS without the DFs of 0.125 and 0.25. In this case, no significant main effect of normalization remained for LBS ($p = .16$) and SBS ($p = .32$). For CR7 even when removing close stimuli (DFs of 0.125 and 0.25) from the analysis a significant effect of normalization ($p < .001$) remained. It can be concluded that for SBS, LBS and TGC normalizing the stimuli only had significant influence on distance estimation for close stimuli which have high DRRs (in this study above 14 dB, see Fig. 3). For these stimuli, large level differences due to the normalization occur. In our study for all these stimuli the level difference exceeded 6 dB(A) (see Fig. 4).

In a next step, we calculated the mean values and the 95 % confidence intervals. The results for the four tested rooms are shown in Fig. 5. Generally, for all rooms and test signals, a good differentiation in estimated distance between the presented distances was achieved. Regarding the loudness-normalized stimuli, distance estimation varied only slightly in comparison to the non-normalized ones. This coincides with the findings of Kolarik [2] who found that both DRR and intensity can serve as robust distance cues.

For further statistical analysis we investigated the influence of the near-field cues and compared the distance estimations for stimuli with near-field HRIRs for close distances to the ones which only far-field HRIRs. In Fig. 6 the respective mean values of estimated distance with 95 % confidence intervals are shown. We performed t-tests with Hochberg-correction which showed only for two of the conditions in CR7 a significant influence (drums, DF = 0.25, loudness normalization, p <.001; guitar, DF = 0.125, loudness normalization, p =.007). In this room, the simulated distances for which near-field HRIRs were considered range from 0.34 m at a DF of 0.125 to 1.35 m at a DF of 0.5. Generally DRR is higher for CR7 than for the other rooms, and thus reverberation probably completely masks the influence of near-field cues in the other rooms. For three out of four tested rooms and for all of the non-normalized stimuli, no significant influence of the near-field HRIRs was noted. It can be stated that if at all, near-field cues only marginally influence distance estimation in reverberant environments.

Finally, we performed t-tests with Hochberg-correction [27] in order to compare the stimuli auralized with the measured BRIRs (originals) to the stimuli generated using synthetic BRIRs with DF of 1. The analysis showed a significant difference for 5 out of 16 conditions at the 0.05 level (CR7: guitar, non-normalized; SBS: guitar, normalized; TGC room: guitar, normalized; guitar, non-normalized; drums, non-normalized). However, even though slight differences in distance estimation for the original BRIR exist, we can assume that the BinRIR algorithm allows to estimate distance comparably to a binaurally measured reference. Similar effects have already been observed in an earlier study [20].

5. Conclusion

We applied the BinRIR algorithm to investigate the contribution of source intensity, DRR, and near-field cues to distance estimation. Based on measured omnidirectional RIRs, we synthesized stimuli for four different rooms at varying source distances. The listening experiment showed that for all investigated rooms, the desired control of the perceived distance could be achieved.

As shown in previous studies, intensity has a strong influence on the results. This manifests for example in the differences that were observed between the normalized and the non-normalized stimuli. Furthermore, the DRR contributes strongly to the estimated distance as well. Even for loudness-normalized stimuli, subjects were able to estimate distance appropriately. In TGC, no significant contribution of the normalization as a main factor could be shown, and for SBS and LBS, normalization only had influence on the results for close distances. Finally, the study delivers information on the relevance of near-field cues for distance estimation. The use of near-field HRIRs created additional variance in the results but no significant trend of their influence could be observed.

The results of this study are relevant for applications in which VAEs are used to present stimuli at different distances, for example in radio dramas or games in which a plausible impression of the environment is desired. Often, a distance control independent of the sound intensity is desired here. Both controlling DRR and intensity allows to change the perceived distance. In future studies it might be interesting to investigate the influence of moving listeners and sound sources on distance estimation. In augmented reality environments specific aspects arise: future research on distance perception might be related to influences of mismatches between real and synthesized objects.

6. Acknowledgment

The research presented in this paper has been carried out in the Research Project MoNRa which was funded by the Federal Ministry of Education and Research in Germany. Support Code: 03FH005I3-MoNRa. We thank the participants of the listening experiments for their patience and commitment.

7. References

[1] Blauert, J., *Spatial Hearing - Revised Edition: The Psychoacoustics of Human Sound Source Localisation*, MIT Press, Cambridge, MA, 1997.

[2] Kolarik, A. J., Moore, B. C. J., Zahorik, P., Cirstea, S., and Pardhan, S., "Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss," *Attention, Perception, & Psychophysics*, 78, pp. 373–395, 2016.

[3] Zahorik, P., Brungart, D. S., and Bronkhorst, A. W., "Auditory distance perception in humans: A summary of past and present research," *Acta Acustica united with Acustica*, 91(3), pp. 409–420, 2005.

[4] Brungart, D. S., Durlach, N. I., and Rabinowitz, W. M., "Auditory localization of nearby sources. II. Localization of a broadband source." *Journal of the Acoustical Society of America*, 106, pp. 1956–1968, 1999.

[5] Shinn-Cunningham, B., "Distance cues for virtual auditory space," *Proceedings of the First IEEE Pacific-Rim Conference on Multimedia*, (December), pp. 227–230, 2000.

[6] Miller, G. A., "Sensitivity to Changes in the Intensity of White Noise and Its Relation to Masking and Loudness," *The Journal of the Acoustical Society of America*, 19(4), pp. 609–619, 1947.

[7] Brungart, D. S. and Rabinowitz, W. M., "Auditory localization of nearby sources. Head-related transfer functions," *Journal of the Acoustical Society of America*, 106(May), pp. 1465–1479, 1999.

[8] Arend, J. M., Neidhardt, A., and Pörschmann, C., "Measurement and Perceptual Evaluation of a Spherical Near-Field HRTF Set," in *Proc. of the 29th Tonmeistertagung - VDT Int. Conv.*, 2016.

[9] Zahorik, P., "Assessing auditory distance perception using virtual acoustics." *The Journal of the Acoustical Society of America*, 111(4), pp. 1832–1846, 2002.

[10] Brungart, D. S. and Simpson, B. D., "Auditory localization of nearby sources in a virtual audio display," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 107–110, 2001.

[11] Schröder, D. and Vorländer, M., "RAVEN: A Real-Time Framework for the Auralization of Interactive Virtual Environments," *Forum Acusticum*, pp. 1541–1546, 2011.

[12] Pörschmann, C. and Wiefling, S., "Perceptual Aspects of Dynamic Binaural Synthesis based on Measured Omnidirectional Room Impulse Responses," in *International Conference on Spatial Audio*, 2015.

[13] Pörschmann, C., Stade, P., and Arend, J. M., "Binauralization of Omnidirectional Room Impulse Responses - Algorithm and Technical Evaluation," in *Proceedings of the DAFx 2017*, pp. 345–352, 2017.

[14] Bernschütz, B., "A Spherical Far Field HRIR / HRTF Compilation of the Neumann KU 100," in *Proc. of the 39th DAGA*, pp. 592–595, 2013.

[15] Bernschütz, B., *Microphone Arrays and Sound Field Decomposition for Dynamic Binaural Recording*, Dissertation, TU Berlin, 2016.

[16] Pörschmann, C. and Zebisch, A., "Psychoakustische Untersuchungen zu synthetischem diffusen Nachhall Psychoacoustic Investigations on synthetically created diffuse Reverberation," in *Proc. of the 27th Tonmeistertagung - VDT Int. Conv.*, pp. 539–550, 2012.

[17] Lindau, A., Kosanke, L., and Weinzierl, S., "Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses," *Journal of the Audio Engineering Society*, 60(11), pp. 887–898, 2012.

[18] Stade, P. and Arend, J. M., "Perceptual Evaluation of Synthetic Late Binaural Reverberation Based on a Parametric Model," in *AES Conference on Headphone Technology*, pp. 1–8, 2016.

[19] Coleman, P., Franck, A., Menzies, D., and Jackson, P. J. B., "Object-based reverberation encoding from first-order Ambisonic RIRs," in *Proceedings of 142nd AES Convention*, pp. 1–10, Berlin, Germany, 2017.

[20] Pörschmann, C. and Stade, P., "Auralizing Listener Position Shifts of Measured Room Impulse Responses," *Proceedings of the DAGA 2016*, pp. 1308–1311, 2016.

[21] Vazquez Giner, A., "Scale - Conducting Psychoacoustic Experiments with Dynamic Binaural Synthesis," *Proceedings of the DAGA2015*, pp. 1128–1130, 2015.

[22] Geier, M., Ahrens, J., and Spors, S., "The soundscape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods," in *Proceedings of 124th Audio Engineering Society Convention 2008*, pp. 179–184, 2008.

[23] Stade, P., Bernschütz, B., and Rühl, M., "A Spatial Audio Impulse Response Compilation Captured at the WDR Broadcast Studios," in *Proc. of the 27th Tonmeistertagung - VDT Int. Conv.*, pp. 551–567, 2012.

[24] Bernschütz, B., Vázquez Giner, A., Pörschmann, C., and Arend, J. M., "Binaural reproduction of plane waves with reduced modal order," *Acta Acustica united with Acustica*, 100(5), pp. 972–983, 2014.

[25] ITU-R BS.1770-4, "Algorithms to measure audio programme loudness and true-peak audio level," 2015.

[26] Greenhouse, S. W. and Geisser, S., "On methods in the analysis of profile data," *Psychometrika*, 24(2), pp. 95–112, 1959.

[27] Hochberg, Y., "A sharper bonferroni procedure for multiple tests of significance," *Biometrika*, 75(4), pp. 800–802, 1988.