DOI: 10.61782/fa.2023.0439

forumacusticum 2023

reproduction of simulated acoustic scenes with limited number of loudspeakers in a reverberant room

Ali Fallah[1*] **Siegfried Gündert**[1] **Christoph Kirsch**[1]
Shun Nakamura[1] **Steven van de Par**[1]
[1] Department of Medical Physics and Acoustics, Cluster of Excellence, Hearing4all, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

ABSTRACT

When reproducing virtual acoustic scenes in a reverberant playback room, the acoustics of the playback room degrades the quality of reproduction. Until now, these scenes can only be optimally rendered on dedicated loudspeaker setups placed in an anechoic room using e.g., Ambisonics or wave field synthesis (WFS). When using virtual scenes in clinical applications, it is desirable to reproduce the sound field with a limited number of loudspeakers in a small reverberant room. Recently we have developed an Acoustic Room Transformation (ART) method based on the Ambisonics that perceptually compensates the reverberation of the playback room by separately capturing and reproducing an optimized version of the direct and reverberant sound fields [1]. Interestingly, when a virtual acoustic scene is created with a room acoustical simulator e.g. [2], the direct and reverberant sound fields are separately available inherently. In this study, the perceptually-based ART method is used to render acoustic environments using only 4 loudspeakers in a reverberant room. A sound-quality evaluation shows that the directional and spectral characteristics of reproduced sound are better preserved when using the ART method compared to standard playback.

Keywords: *Virtual Acoustics, Room Simulation, Acoustic Room Transformation (ART), Spatial Rendering in Room*

Corresponding author: ali.fallah@uol.de.

1. INTRODUCTION

Room simulation are used extensively in virtual acoustics to create immersive audio experiences, particularly in fields such as music production, gaming, and virtual reality. In many room simulations methods, it is possible to render a simulated room on both headphones and loudspeakers. The CAVE system [3] is an Automatic Virtual Environment (AVE) designed to provide a highly immersive audio-visual experience. This system consists of a room-sized cube with walls made of rear-projection screens, high-quality audio speakers, and motion tracking sensors. The Virtual Reality (VR) system at RWTH Aachen University [4], is a comprehensive platform for research and development in the field of VR. The system is designed to support a wide range of VR applications, including immersive visualization, simulation, and interaction. The Simulated Open-Field Environment (SOFE) [5] is another system that simulates and reproduces audio–visual environments in a laboratory setting. This system consists of a visual display, a loudspeaker array and a head-tracking device that allows for the creation of a 3D sound field. When sound is reproduced using an array of loudspeakers, it must be perceived by the listener in a way that is similar to what is heard in the real recorded world. This allows for more ecologically valid experiments and helps to ensure that the results obtained are applicable to real-world scenarios [6]. The immersion of reproduction techniques for room auralizations which aims to improve the audio experience in virtual environments are investigated in [7]. In the most of loudspeaker-based rendering methods such as e.g. [4, 5], the reproduction is performed in an anechoic room. But, in clinical application that aims evaluation of speech intelligibility, there is usually no access to anechoic rooms and also large loudspeaker setups. Therefore, it is interesting to reproduce a simulated room on a limited number of loudspeakers in a normal reverberant room.

Audio playback in a room over loudspeakers results in a “Room-In-Room” (RinR) response that can increase the reverberation time, strongly modify the perceived coloration, change the temporal envelope of the early reflections, increase in spectral modulation strength, and decrease speech intelligibility that is finally resulting in an unnatural and impaired listening experience [8]. The Acoustic Room Transformer (ART) method that we propose here aims to transform the acoustic properties of a given space to match those of a desired target space resulting in an enhanced spatial audio playback. We previously used this perceptually-based ART method to render real-room recordings with a loudspeaker array in a reverberant room [1]. In this ART method, direct and reverb parts of a recorded source are separated, compensated and rendered in a reverb playback room, where in the process of compensation, the detrimental effect of the reproduction environment is reduced. For room simulation methods, the direct and reverb parts are separately available and therefore it is easier to apply the ART method. In this paper, two rooms that are simulated by RAZR [9] are rendered using the ART method over four loudspeakers in a normal room and perceptually evaluated.

2. METHOD

The perceptually-based ART system, used for rendering of a simulated room is depicted in Figure 1. The RAZR is used as a room simulator that provides a simulated post-spatialized binaural room impulse response (BRIR), but also simulated direct and reverb sources that are mapped on an optional loudspeaker setup using Vector-Base Amplitude Panning (VBAP) approach [9]. The mapped direct and reverb parts are perceptually compensated and rendered over four loudspeakers in a reverberant playback room. The KEMAR dummy head, placed in the reproduction room, is compared with the simulated post spatialized BRIR using the same KEMAR dummy head in order to obtain the compensation filters depicted in Figure 1. Two perceptually-based criteria used to obtain the compensation filters are the energy and the Interaural Coherence (IC). The azimuth positions of loudspeakers with respect to the dummy head are 45°, 135°, 225° and 315° degrees. A simulated post-spatialized reference signal can be divided into direct and reverberation parts as separate signals:

$$BRIR_{ref}[n] = BRIR_{ref,dir}[n] + BRIR_{ref,rev}[n]. \tag{1}$$

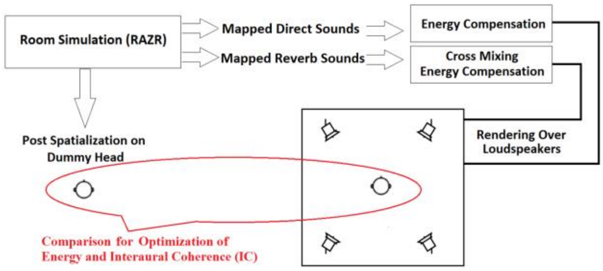

Figure 1. The audio reproduction system: The outputs of RAZR for direct and reverb parts of simulated RIRs are perceptually compensated and rendered over four loudspeakers in a reverb room. The dummy head in the reproduction room is compared with the simulated post spatialized dummy head in RAZR to obtain the compensation filters considering the energy and the interaural coherence (IC) criteria.

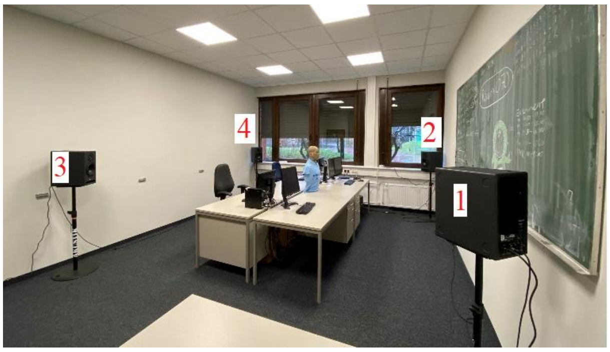

Figure 2. The setup of reproduction room that is including four loudspeakers in azimuth positions of 45°, 135°, 225° and 315° degrees.

From here, BRIR in all equations can be considered as a signal of left or right ear. Note that the separation time for direct and reverb parts of $BRIR_{ref}$ denoted by T_{ref} plays an important role in our ART method. Because in the reproduction setup there is no loudspeaker outside the horizontal plane, the elevation angles in the mapping on loudspeakers are discarded. It is also assumed that the direct signal arrives from front and horizontal plane. Therefore, the direct sound output of RAZR is mapped on loudspeakers number one and two in Figure 2. The mapped direct signal on loudspeakers one and two are depicted by $L_{1,dir}[n]$ and $L_{2,dir}[n]$. For optimization, the

BRIRs of loudspeakers ($BRIR_{play,l}[n]$) in the reproduction room are used, where $1 \le l \le 4$ is the loudspeaker number. This BRIR of playback room similar to the post spatialized reference BRIR in Eq (1) includes the direct and reverb parts:

$$BRIR^{(i)}{}_{play,l}[n] = BRIR^{(i)}{}_{play,l,dir}[n] + BRIR^{(i)}{}_{play,l,rev}[n],\ 1 \le l \le 4. \tag{2}$$

The direct played back signal has the following binaural room-in-room impulse response ($BRinRIR_{dir}$):

$$BRinRIR_{dir}[n] = L_{1,dir}[n] * BRIR_{play,1,dir}[n] + L_{2,dir}[n] * BRIR_{play,2,dir}[n] \tag{3}$$

The direct and reverb sounds are separately filtered by a 4th-order Gammatone filterbank [10]. The Auditory Transfer Function ($ATF^{(i)}$) defined as the energy of a signal at the output of ith filter of Gammatone filterbank is compensated to be similar to that of the reference signal in Eq (1). The first step is the compensation of direct sound ($BRinRIR_{dir}$). For the energy equalization of direct sound, the Gammatone filterbank gains ($g_{dir}^{(i)}$) are used:

$$ATF^{(i)}\left\{BRIR_{ref,dir}[n]\right\} = \left(g_{dir}^{(i)}\right)^2 ATF^{(i)}\left\{BRinRIR_{dir}[n]\right\}. \tag{4}$$

where $g_{dir}^{(i)}$ is the gain of loudspeakers number one and number two in frequency band number i. A method proposed in [11] is used for gain optimization taking into account the effect of overlapping filters such as the Gammatone. It must be mentioned that two different gains are obtained by considering left and right ears and the average value of these gains are finally used as the gains of the loudspeakers used for direct sound. The new energy-compensated (Eco) signals of loudspeakers for the direct sound are:

$$L^{(i)}{}_{1,dir,ECo}[n] = g_{dir}^{(i)} L^{(i)}{}_{1,dir}[n],\quad L^{(i)}{}_{2,dir,ECo}[n] = g_{dir}^{(i)} L^{(i)}{}_{2,dir}[n]. \tag{5}$$

After using this energy-compensated signals for the loudspeakers number one and two in Eq (3), the compensated $BRinRIR$ of the direct sound ($BRinRIR_{dir,C}$) is obtained:

$$BRinRIR_{dir,C}[n] = L^{(i)}{}_{1,dir,ECo}[n] * BRIR_{play,1,dir}[n] + L^{(i)}{}_{2,dir,ECo}[n] * BRIR_{play,2,dir}[n]. \tag{6}$$

After the compensation of direct sound, the reverb sound is compensated. The mapped signals onto the loudspeakers using VBAP for the reverb sound are depicted by $L_{l,rev}[n], 1 \le l \le 4$. The $L_{l,rev}[n]$ includes a summation of all of the reverb sounds mapped on the loudspeaker number l. The optimization of reverb sound energy is being jointly performed with the interaural coherence (IC) optimization of the whole reproduced signal. To control the IC, the signals of front and back loudspeakers are separately cross mixed in different frequency bands. The new cross-mixed (CM) signals of all loudspeakers are obtained by cross mixing the signals of two front (L_1 and L_2) and two back loudspeakers (L_3 and L_4):

$$\begin{aligned} L_{1,rev,CM}^{(i)}[n] &= L_{1,rev}^{(i)}[n] - \alpha_i \times L_{2,rev}^{(i)}[n], \\ L_{2,rev,CM}^{(i)}[n] &= L_{2,rev}^{(i)}[n] - \alpha_i \times L_{1,rev}^{(i)}[n], \\ L_{3,rev,CM}^{(i)}[n] &= L_{3,rev}^{(i)}[n] - \beta_i \times L_{4,ev}^{(i)}[n], \\ L_{4,ev,CM}^{(i)}[n] &= L_{4,rev}^{(i)}[n] - \beta_i \times L_{3,rev}^{(i)}[n]. \end{aligned} \tag{7}$$

The coefficients α_i and β_i are obtained by a search algorithm over a range of possible values. A reproduced $BRinRIR$ for reverb sound in each frequency band using these cross-mixed signals is:

$$BRinRIR_{rev}^{(i)}[n] = \sum_{l=1}^{4} L_{l,rev,CM}^{(i)}[n] * BRIR_{play,l}^{(i)}[n] + \sum_{l=1}^{2} L_{l,dir,ECo}^{(i)}[n] * BRIR_{play,l,rev}^{(i)}[n], \tag{8}$$

where $BRIR_{play,l,rev}[n]$ is the reverberant part of BRIR in the reproduction room in Eq (2). The second part of right side of Eq (8) is included because the compensated direct sounds on the front loudspeakers ($L_{l,dir,ECo}^{(i)}$) after being used in Eq (6) also produce reverberant sounds. The separation time of the direct and reverb parts of Eq (2) shown by T_{play} is also very important in the optimization procedure. A correct selection of T_{ref} related to Eq (1) and T_{play} determines the quality of the reproduced summed direct and reverb sounds. This will be discussed in the next section. The energy compensation for the reverberant part is performed similar to the direct sound:

$$ATF^{(i)}\left\{BRIR_{ref,rev}[n]\right\} = \left(g_{rev}^{(i)}\right)^2 ATF^{(i)}\left\{BRinRIR_{play,rev}[n]\right\}. \tag{9}$$

After the energy compensation of reverb sound, the Eq (8) is modified to:

$$BRinRIR_{rev}^{(i)}[n]=\sum_{l=1}^{4}L_{l,rev,CM.ECo}^{(i)}[n]*BRIR_{play,l}^{(i)}[n] +\sum_{l=1}^{2}L_{l,dir,ECo}^{(i)}[n]*BRIR_{play,l,rev}^{(i)}[n], \quad (10)$$

in which $L_{l,rev,CM.ECo}^{(i)}[n]=g_{rev}^{(i)}L_{l,rev,CM}^{(i)}[n]$ is called cross-mixed and energy-compensated (CM.ECo) reverb signal. By combining Eq (6) and Eq (10), the final compensated BRinRIR is obtained:

$$BRinRIR_{Co}^{(i)}[n]= BRinRIR_{dir,Co}^{(i)}[n]+BRinRIR_{rev,Co}^{(i)}[n]. \quad (11)$$

The Interaural Cross Correlation (IACC) and Interaural Coherence (IC) between left and right ears after cross mixing and energy-compensation are evaluated in this step:

$$IACC^{(i)}[q]= \frac{\sum_{m=-\infty}^{m=\infty}BRinRIR_{left,Co}^{(i)}[m]BRinRIR_{right,Co}^{(i)}[m+q]}{\sqrt{\sum_{m=-\infty}^{m=\infty}\left(BRinRIR_{left,Co}^{(i)}[m]\right)^2}\sqrt{\sum_{m=-\infty}^{m=\infty}\left(BRinRIR_{right,Co}^{(i)}[m]\right)^2}}, \quad (12)$$

$$IC^{(i)}=\max(IACC^{(i)}).$$

For each frequency band, this $IC^{(i)}$ is compared with the IC between $BRIR_{ref,left}[n]$ and $BRIR_{ref,right}[n]$ related to the reference signal in Eq (1). The error value in each frequency band is $IC_{error}^{(i)}=\left|IC^{(i)}-IC_{ref}^{(i)}\right|$. To find the best values for α^i and β^i, a searching procedure repeatedly performs an optimization using equations (7-12). The α^i and β^i related to a minimum value of the $IC_{error}^{(i)}$ with the corresponded $g_{rev}^{(i)}$ are respectively the final cross mixing coefficients and filter gains of the search procedure in one frequency band. The signals of loudspeakers for reverb sound in each band are obtained using the optimized cross mixing coefficient and energy compensating gains:

$$\begin{aligned}
L_{1,rev,CM.ECo}^{(i)}[n]&=g_{rev}^{(i)}\left(L_{1,rev}^{(i)}[n]-\alpha_i\times L_{2,rev}^{(i)}[n]\right),\\
L_{2,rev,CM.ECo}^{(i)}[n]&=g_{rev}^{(i)}\left(L_{2,rev}^{(i)}[n]-\alpha_i\times L_{1,rev}^{(i)}[n]\right),\\
L_{3,rev,CM.ECo}^{(i)}[n]&=g_{rev}^{(i)}\left(L_{3,rev}^{(i)}[n]-\beta_i\times L_{4,ev}^{(i)}[n]\right),\\
L_{4,rev,CM.ECo}^{(i)}[n]&=g_{rev}^{(i)}\left(L_{4,rev}^{(i)}[n]-\beta_i\times L_{3,rev}^{(i)}[n]\right).
\end{aligned} \quad (13)$$

The final compensated signals of loudspeakers are obtained by Gammatone synthesizing of compensated direct and reverb sounds:

$$L_{l,Co}[n]= \mathrm{Re}\,synthesis\left\{\begin{matrix}L_{l,dir,CM.ECo}^{(i)}[n]\\ +L_{l,rev,CM.ECo}^{(i)}[n]\end{matrix}\right\}, 1\le l\le 4. \quad (14)$$

3. EVALUATION OF THE ART SYSTEM

Two simulated rooms in RAZR, referred to as room 1 and room 2 are used for evaluation of the proposed method. The reverberation times of room 1 and room 2 are $T_{60,room1}$=0.65 and $T_{60,room2}$=1.15 second respectively and the reverberation time of playback room is about $T_{60,playback}$=0.70 second. For the optimization, the separation times of direct and reverb sounds according to the Equations (1) and (2) are selected empirically T_{ref}=10ms, T_{play}=16ms for room 1 and T_{ref}=10ms, T_{play}=20ms for room 2. These separation times allow to control the reverberation time and also the direct-to-reverb ratio in the reproduction room.

The results of compensation to render room 1 are shown in Figure 3. Two upper panels in Figure 3 show the ATFs of reproduced BRIRs for left and right ears in which the energy of compensated BRIR is compared with that of the reference, uncompensated and also compensated direct cases. For these plots, the root-mean-squared (RMS) values of all BRIRs are normalized for the right ear that has higher energy. Note that for the uncompensated case, the reverberation of playback room is leading to an increase of energy specifically in the low frequencies. The compensated BRIR in comparison to uncompensated case fits better to the energy curve of the reference signal specifically in the middle and high frequencies. In the lower frequency bands, where the reverberation is dominant (for example bands number three and four), a complete compensation is not achieved. This is because of very similar reverberation times of the simulated and playback rooms in which according to Eq (8), the direct sound reproduces a reverb field in the playback room that is not controllable anymore by adding the reverb sound of simulated room. Because most of the reverb energy is in the low frequency range, the compensated direct BRIR in the two upper panels shows a low amount of energy in low frequencies in comparison with the high frequencies. After addition of the compensated reverb field, these low-energy bands in low frequencies are elevated compared to that of reference signal. In the lower-left panel of Figure 3, the ICs of uncompensated and uncompensated BRIRs are compared to that of the reference signal in the frequency bands up to number 14. These bands are perceptually most important for IC because for frequencies below 1500 Hz, the auditory system is most sensitive to the changes in IC. Most of Gammatone filter bank outputs show an improvement for ICs, specifically in frequency bands near to 1500 Hz. For the ICs, in some frequency bands, there are no improvements, similar to the ATF curves. This is again related to the very similar reverberation times of simulated and playback rooms. The lower-left panel show the Schroeder's energy decay curves (EDCs) [12] of the reference, compensated and Compensated BRIRs.

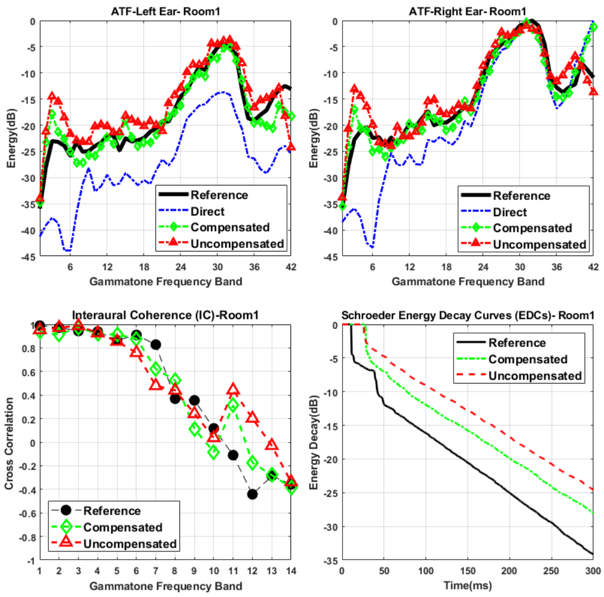

Figure 3. The results of compensation and comparison with the uncompensated BRIRs in room 1. In two upper panels, the ATF of BRIRs of reference, uncompensated and compensated BRIRs for the right and left ears shown. There are improvements of energy in both ears. In lower-left panel, the ICs of compensated and uncompensated and their comparisons with that of reference BRIR for perceptually important frequencies below 1500 Hz are depicted. The lower-right panel show the Schroeder' energy decay curves (EDCs) of BRIRs. There are improvements in EDC and subsequently in reverberation time of compensated case.

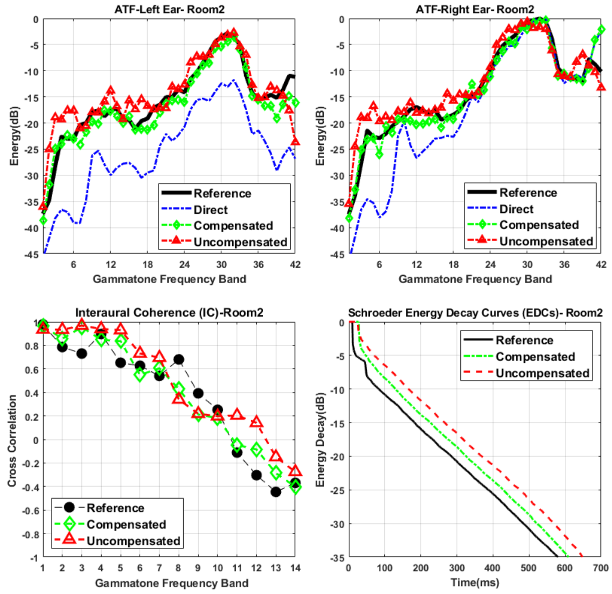

Figure 4. Similar to the Figure 3, the ATFs, ICs and EDCs are plotted for room 2. The ATFs and ICs of compensated BRIR show a good match to that of the reference case also in the low frequency bands. Also in lower-right panel, the improvement of EDC and reverberation time of compensated BRIR in comparison to that of uncompensated case is depicted.

As can be noted, our approach can also improve the reverberation time. The EDC of compensated BRIR, in comparison to the uncompensated case is improved and is closer to that of the reference case. The ATF, ICs and EDCs related to room 2 that now has a clearly higher reverberation time than the playback room are depicted in Figure 3. In comparison to room 1, there is a good match between the compensated and the reference BRIRs for both ATFs and ICs for whole frequency range. Specifically, there is a better compensation in room 2 in low frequencies compared to room 1. This is clear in two upper panels of Figure 4 in which a good matching between energy of compensated BRIR and that of reference case is seen. This good matching is because of lower reverberation time of reproduction room in comparison to that of simulated room 2. In this situation, the control of reverb field in the reproduction room is possible by adding more reverb sound of the simulated room. Moreover, in lower-left panel, the improvement of ICs for the compensated BRIR in comparison to

uncompensated case is depicted across all frequency bands. Also, similar to room 1 case, the EDCs and reverberation time of compensated BRIR are improved, as is depicted in the lower-right panel of this Figure 4.
Overall, the results of the presented optimizations for rooms 1 and 2 showed the importance of the reverberation time of the simulated rooms. If the reverberation time of simulated room is lower than that of playback room, a complete compensation for ATF and IC for whole frequency band is possible. Otherwise, for close reverberation times, the compensation for some frequency bands can be difficult or impossible.
For evaluation of quality, a multi-stimulus scaling experiment was performed with 16 listeners over headphones similar to the MUSHRA test [13]. The signals used in this experiment are an ideally simulated post-spatialized reference signal with RAZR also used as hidden reference (Ref), A compensated direct signal played from front loudspeakers (Direct), a reproduced direct and reverb signals in the playback room without compensation (Uncomp), the compensated signal produced by our proposed ART method (Comp) and the BRIR measured by a real KEMAR in the real two rooms (KEMAR). The real BRIRs of real KEMAR were previously used in RAZR to support the simulation of the two rooms. For each room, one male voice and five instruments including guitar, clarinet, piano, snare drum and trumpet were used. In the listing test, subjects were instructed that the hidden reference must be scored 100%. Figure 5 shows the subjective variations for averaged scores of all instruments for all subjects. Here, the median values are shown together with the 25% and 75% quantiles and outliers across 16 subjects. The mean and standard errors of data in Figure 5 are depicted in Figure 6. The reproduction of direct sound only leads to a strong perceived signal degradation compared to the reference signal. The uncompensated case, which includes direct and reverb sound, shows already strong improvement in comparison to only direct reproduction but still has a degraded quality in comparison to the reference signals. For both rooms, the compensated audios are scored higher than uncompensated cases. For the room 1, the measured KEMAR signals show high similarities to the reference in comparison to the compensated reproduced sounds. But in the room 2, the similarity of the compensated reproduced sounds to the reference is higher than that of measured KEMAR. The main reason for this is the higher reverberation time in the simulated room 2 compared to room 1.

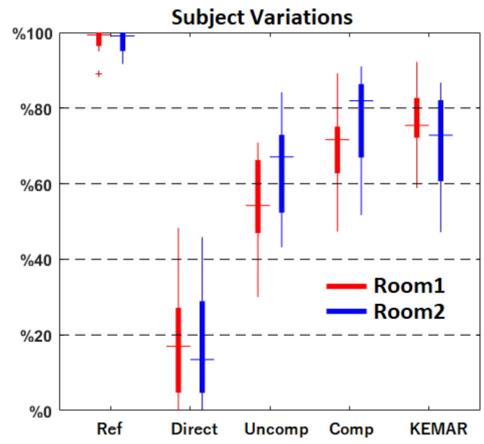

Figure 5. Boxplots of subject variations for listeners' scorings averaged over all instruments. Median values are shown together with the 25% and 75% quantiles and outliers across 16 subjects. For both rooms there are improvements for the compensated (Comp) signals in comparison to the uncompensated (Uncomp) cases. Also, for room 2 that is more reverberant, the scores are higher than that of the original real room measurement (KEMAR).

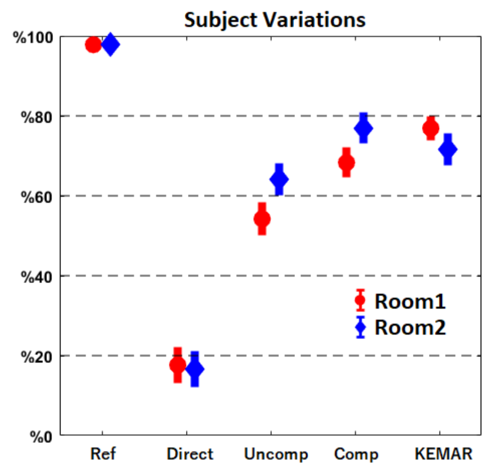

Figure 6. The mean and standard errors across the subjects for the data in Figure 3. The red and blue symbols illustrate the results for room 1 and room 2, respectively. In both rooms there is 15% improvement in ART rendering in comparison to the uncompensated case.

When the reverberation in the playback room is lower than that in the simulated room, it is easier to add extra reverb sound in the playback room to control the sound field. However, compensating for the energy difference between the two rooms becomes very challenging, and in some cases impossible, when the reverberation time in the reproduction room is greater than that of the playback room. It is nevertheless shown that despite the similarity of the reverberation times of the simulated room and the reproduction room 1, a good quality rendering is obtained.

The mean and standard errors across each instrument for room 1 and room 2 are depicted respectively in Figure 7 and Figure 8. For room 1 in Figure 7, all instruments except clarinet apparently show enhancement for compensated cases in comparison to the uncompensated signals. The similarities of measured KEMAR to the references except piano and trumpet are higher that of the compensated cases. For room 2, according to the Figure 8, all instruments except piano show improvements in comparison to the uncompensated case. The score of clarinet, for the compensated case is only a bit better than the uncompensated case. Also, the scores of compensated audio examples, for all instruments except trumpet, are better than the measured KEMAR. For the Clarinet, small differences between the compensated and uncompensated cases are seen in both rooms. The lack of improvement in compensating for reverberation in clarinet recordings may be primarily attributed to the instrument's spectral characteristics.

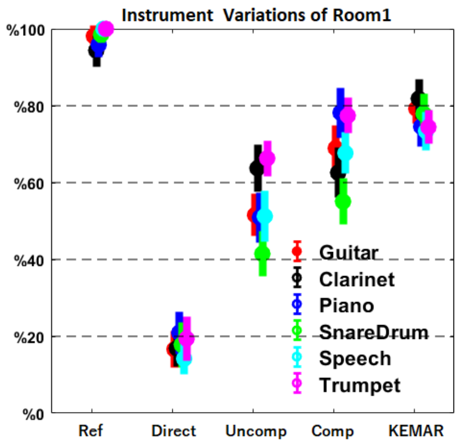

Figure 7. The mean and standard errors across each instrument in room 1. Except clarinet, there are apparent improvements for compensated signals in comparison to the uncompensated cases.

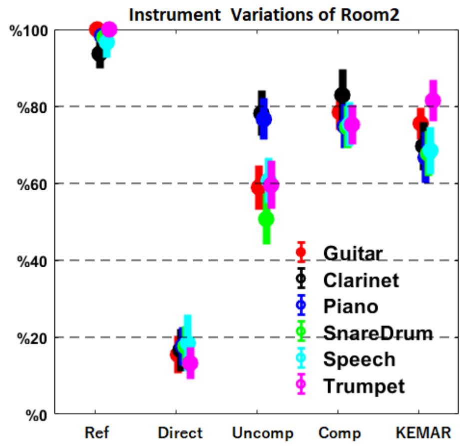

Figure 8. The panel is similar to Figure 5 but for room 2. Except for clarinet and piano, there are clear improvements for compensated signals in comparison to the uncompensated cases.

The clarinet produces most of its energy in the high-frequency range, which is less susceptible to the benefits of reverberation compensation techniques. The same is seen for piano but only in room 2.

4. SUMMARY

For spatial reproduction of simulated scenes such as created with a tool like RAZR [2, 9], a method using a low number of loudspeakers is proposed that is perceptually compensating for the acoustics of a playback room. The simulated direct and reverb sounds are filtered and played back using the VBAP method on four loudspeakers in a reverberant room. In this approach, instead of ideally reconstructing the simulated sound field, the directional cues, energy and interaural coherence (IC) of sounds are reproduced as accurately as possible matching the simulated reference signal. The energy optimizations are separately performed for direct and cross-mixed reverb sounds. The cross-mixed reverb sounds on all loudspeakers can control the ICs of the rendered signals. Moreover, the energy decay curve of the reproduced signal is improved. It is also interesting to note that the reflections related to elevation angles are not represented at all due to the VBAP mapping, because all four loudspeakers are in the horizontal plane. The elevation information will be reproduced to some extend using the reverberation of the playback room. The results of the proposed room compensation, showed that the reproduced energy and IC for a compensated BRIR have a good matching to that of a reference case. If the reverberation time of a simulated room is larger than that of a playback room, this compensation is obtained in the whole frequency range. Otherwise, for the case that these reverberation times are close to each other, the compensation can't be completely performed, specifically in low frequencies in which the reverberation is dominant. In this case, the direct sound, because of the reverberation of a playback room, could result in producing an excessive amount of reverb field, that cannot be control effectively. Therefore, as a limitation of the ART system, it is important to use only simulated rooms with reverberation times lower than that of playback room to enable effective control of the reverb energy. The results of listening test also showed an apparent quality improvement of compensated signals in comparison to the uncompensated cases for speech and most of instruments. For the current loudspeaker setup, the direct sound is limited to the horizontal plane. Possibly the ART method can be improved by adding more loudspeakers in the horizontal plane and one ceiling loudspeaker.

5. ACKNOWLEDGMENTS

The authors of this paper would like to thank the German Research Foundation (DFG) for funding this work (Project-ID: 352015383–SFB 1330 C2) as part of the HAPPAA.

6. REFERENCES

[1] A. Fallah and S. Par, "A new approach for rendering Ambisonics recordings on loudspeakers in a reverberant environment," in Forum Acusticum, pp. 2071-2077, 2020.

[2] T. Wendt, S. Van De Par, and S. D. Ewert, "A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation," Journal of the Audio Engineering Society, vol. 62, no. 11, pp. 748-766, 2014.

[3] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart, "The CAVE: audio visual experience automatic virtual environment," Communications of the ACM, vol. 35, no. 6, pp. 64-73, 1992.

[4] D. Schröder, F. Wefers, S. Pelzer, D. Rausch, M. Vorländer, and T. Kuhlen, "Virtual reality system at RWTH Aachen University," in Proceedings of the international symposium on room acoustics (ISRA), Melbourne, Australia, 2010.

[5] B. U. Seeber, S. Kerber, and E. R. Hafter, "A system to simulate and reproduce audio–visual environments for spatial hearing research," Hearing research, vol. 260, no. 1-2, pp. 1-10, 2010.

[6] S. E. Favrot, A loudspeaker-based room auralization system for auditory research. Technical University of Denmark, 2010.

[7] L. Aspöck, A. Colsman, M. Kohnen, and M. Vorländer, "Investigating the immersion of reproduction techniques for room auralizations," Jahrestagung für Akustik, DAGA, vol. 2016, pp. 565-568, 2016.

[8] A. Haeussler and S. van de Par, "Crispness, speech intelligibility, and coloration of reverberant recordings played back in another reverberant room (Room-In-Room)," The Journal of the Acoustical Society of America, vol. 145, no. 2, pp. 931-944, 2019.

[9] C. Kirsch, J. Poppitz, T. Wendt, S. van de Par, and S. D. Ewert, "Spatial resolution of late reverberation in virtual acoustic environments," Trends in Hearing, vol. 25, p. 23312165211054924, 2021.

[10] V. Hohmann, "Frequency analysis and synthesis using a Gammatone filterbank," Acta Acustica united with Acustica, vol. 88, no. 3, pp. 433-442, 2002.

[11] S. Van De Par, V. Kot, and N. Van Schijndel, "Scalable noise coder for parametric sound coding," in 118th Convention of the Audio Engineering Society, J. Audio Eng. Soc.(Abstracts), 2005, vol. 53, p. 699.

[12] M. R. Schroeder, "New method of measuring reverberation time," The Journal of the Acoustical Society of America, vol. 37, no. 6, pp. 1187-1188, 1965.

[13] I. BS, "1534-1, Method for the subjective assessment of intermediate quality level of coding systems," International Telecommunications Union, Geneva, Switzerland, vol. 14, 2003.