



**HAL**  
open science

# A new approach for rendering Ambisonics recordings on loudspeakers in a reverberant environment

Ali Fallah, Steven van De Par

► **To cite this version:**

Ali Fallah, Steven van De Par. A new approach for rendering Ambisonics recordings on loudspeakers in a reverberant environment. Forum Acusticum, Dec 2020, Lyon, France. pp.2071-2077, 10.48465/fa.2020.0440 . hal-03235352

**HAL Id: hal-03235352**

**<https://hal.science/hal-03235352>**

Submitted on 27 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A new approach for rendering Ambisonics recordings on loudspeakers in a reverberant environment.

Ali Fallah<sup>1</sup>

Steven van de Par<sup>1</sup>

<sup>1</sup> Hearing4all, Acoustics group, CvO University Oldenburg, Germany

ali.fallah@uni-oldenburg.de

## ABSTRACT

It is desirable to reproduce a recorded sound using loudspeakers in a way that a listener has the impression of hearing the original sound. In practice, the distortions related to the reverberation of the reproduction room and non-ideal responses of loudspeakers and position misalignments are inevitable. We propose a method that records, compensates and reproduces the direct and reverb parts of a signal separately. The recording is performed by a spherical microphone array and the spherical harmonics (SHs) coefficients are extracted. After the beamforming which enhances the level of the direct sound, the reverb sound is extracted by subtraction of spatialized direct sound from the original SH coefficients. The rendering of the direct and reverberant sound field is spectrally adjusted based on perceptually optimization of room-in-room (RinR) impulse responses in each auditory frequency band. The direct and reverb parts are filtered in a way that the desired energy and Interaural Coherence (IC) of the recorded sound field are obtained for both the direct and the reverberant sound component. This method provides a perceptually good rendering of the original recorded sound field and compensates for the acoustics of the reproduction environment.

## 1. INTRODUCTION

Theoretically, reproduction of a recorded sound field is possible by wave-field synthesis (WFS) [1] and Ambisonics [2]. However, in both cases the reproduction room must have boundaries without reflection. In practice, this is not possible. Moreover, the non-flat frequency response of loudspeakers and their misalignments are additional causes for non-ideal reproduction. Compensation approaches have been proposed to enhance the sound quality. The main approach in the Ambisonics domain is the modification of ideal weights determined for anechoic reproduction [3]. Recently a modified version of this method was proposed to reduce the computational load [4]. In practice, available recording microphones support 4<sup>th</sup> order Ambisonics. For this order, the area that is reconstructed accurately is limited for high frequencies and is placed in the central area of the loudspeaker array. Complementary to these approaches, there are methods to reproduce the original sound considering perceptual criteria. In the Directional Audio Coding (Dirac) [5], the direct and reverb sounds are transferred in different processing paths. The direct sound is reproduced by

Vector Base Amplitude Panning (VBAP) method [6] decoding the directional cues and the reverb sound is reproduced by multiple loudspeakers using decorrelation methods incorporating InterAural Cross Correlation (IACC) information. In this method, the reverberation of the reproduction room is not considered. A modified version of this system using a microphone array is introduced that can handle more than one simultaneous sound sources [7]. A method for recording and reproduction of direct and reverb sounds by perceptually compensation of the reproduction room was proposed in [8]. Here, three distributed microphones are used to record direct and reverb sounds and four loudspeakers are used for reproduction. This method separately equalizes direct and reverb sound energies in each frequency bands by equalization of RinR impulse responses. The optimization procedure also compensates Interaural Coherence (IC) of left and right ear signals for the reverberant sound field.

In this study the perceptually compensation idea proposed in [8] is combined with Ambisonics. The recording and reproduction are performed by a standard Ambisonic setup and direct and reverb parts of a recorded sound are perceptually compensated. This paper is organized in the following way. In section 2, the basics of Ambisonics recording, beamforming, separation of direct and reverb sounds and standard Ambisonic reproduction are explained. Then in section 3, the signals used for optimization are described. After that, the proposed optimization procedure is explained in section 4. In section 5, results of simulations and an evaluation of the algorithm are presented. Finally, we have a summary in section 6.

## 2. SIGNAL MODEL, RECORDING AND REPRODUCTION

### 2.1 SHs recording

The spherical harmonics (SHs) definition of order  $n$  and degree  $m$  [9] is:

$$Y_n^m(\theta, \varphi) \equiv \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta) e^{im\varphi}, \quad (1)$$

where  $\theta$  and  $\varphi$  are elevation and azimuth angles, respectively, and  $P_n^m$  is the associated Legendre polynomials of order  $n$  and degree  $m$ . For a point source in position  $\mathbf{r}_s$  with amplitude  $S$  and wavenumber  $k$ , the SHs expansion is:

$$S \frac{e^{-ik\|\mathbf{r}-\mathbf{r}_s\|}}{\|\mathbf{r}-\mathbf{r}_s\|} = -4\pi i S k \sum_{n=0}^{\infty} \sum_{m=-n}^n h_n^{(2)}(kr_s) j_n(kr) Y_n^m(\theta_s, \varphi_s) Y_n^m(\theta, \varphi), \quad (2)$$

where  $j_n(kr)$  is the  $n$  order spherical Bessel function of the first kind and  $h_n^{(2)}$  is the spherical Hankel function of the second kind. For a recorded point source in a room, the summation of direct and reverb sound pressure  $P$  in term of SHs expansion is:

$$P(r, \theta, \varphi, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n A_n^m(k) b_n(kr) Y_n^m(\theta, \varphi), \quad (3)$$

where  $A_n^m(k)$  are SHs coefficients and  $b_n(kr)$  depends on the scattering of recording spherical array microphone with radius  $r_e$ :

$$b_n(kr) = \begin{cases} j_n(kr) - \frac{j_n'(kr_e)}{h_n^{(2)'(kr_e)}} h_n^{(2)}(kr), & \text{for a rigid sphere} \\ j_n(kr) & \text{for an open sphere} \end{cases}. \quad (4)$$

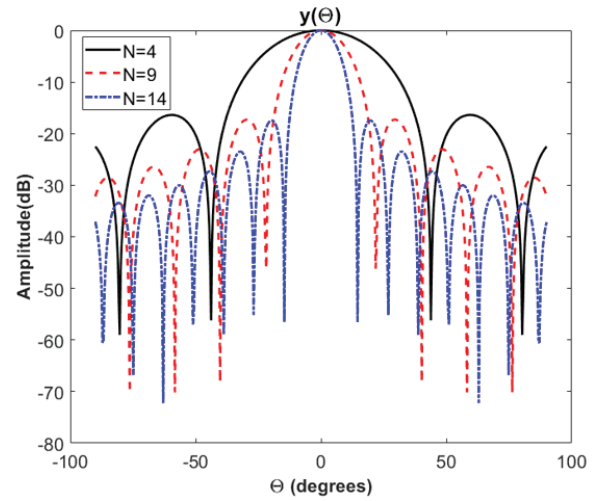
The spatial information of all virtual point sources are summed in one spherical harmonic coefficient  $A_n^m(k)$ . In practice, the summation for  $n$  in Eq (3) is truncated to a finite order  $N$ . The SHs recording is performed by a microphone array with radius  $r_e$  and  $Q$  elements and element weights  $\alpha_q$ :

$$A_n^m(k) = \frac{\sum_{q=1}^Q \alpha_q P(r_e, \theta_q, \varphi_q, k) Y_n^m(\theta_q, \varphi_q)^*}{b_n(kr_e)} \quad (5)$$

## 2.2 Direct and reverb sounds capturing

The spherical microphone array is used to capture the direct sound component by creating a beamformer focussed on the sound source. We use an axis-symmetric beamformer [9] and Dolph-Chebyshev weighting [10]  $d_{cheby, n}$  for a finite order  $N$ :

$$S_{BF}(k, r_a) = \frac{\sum_{n=0}^N \sum_{m=-n}^n \frac{d_{cheby, n}}{C_{ref}} Y_n^m(\theta_{DOA}, \varphi_{DOA}) A_n^m(k)}{\sum_{n=0}^N \frac{2n+1}{4\pi} d_{cheby, n}} \quad (6)$$



**Figure 1.** The array beam pattern for the axis-symmetric beamformer in equation (10) for  $N=4,9$  and  $14$ .

where

$$C_{ref} = -4\pi i k h_n^{(2)}(kr_{ref}), \quad (7)$$

are the reference weights set for a direct sound source with an approximate distance  $r_{ref}$  from the recording microphone. The near field model is considered, because for the lower frequencies the assumption of the plane wave model is not valid. The direction of arrival (DOA) can be predetermined or estimated. To extract the reverb signal, first the direct sound is spatialized:

$$S_{n\ BF}^m(k, r_a) = S_{BF}(k, r_a) \left[ Y_n^m(\theta_{DOA}, \varphi_{DOA}) \right]^*, \quad (8)$$

and then is reduced from original SHs coefficients:

$$S_{n\ Rev}^m(k, r_a) = A_n^m(k) - S_{n\ BF}^m(k, r_a). \quad (9)$$

In this computation, it is assumed that only one source is active within each time frame. The array beam pattern for the axis-symmetric beamformer is:

$$y(\Theta) = \sum_{n=0}^N d_n \sum_{m=-n}^n Y_n^m(\theta_{DOA}, \varphi_{DOA}) Y_n^m(\theta, \varphi)^* = \sum_{n=0}^N d_n \frac{2n+1}{4\pi} P_n(\cos \Theta). \quad (10)$$

where  $\Theta$  is the sold angel between  $(\theta_{DOA}, \varphi_{DOA})$  and  $(\theta, \varphi)$ . The beam pattern in Eq (10) after normalization is plotted in Figure 1 for the uniform weighting ( $d_n = 1$ ) and three different orders. The highest order of Ambisonics that is usually used in recording microphones is 4. According to Figure 1 and equation (9) for this order, most of direct sound energy will be suppressed even if a small error in estimation of DOA exists.

### 2.3 Direct and reverb sounds reproduction

The direct sound obtained from the beamformer output  $S_{BF}(k, r_a)$  is reproduced using VBAP to auralize it at the positions indicated by  $(\theta_{DOA}, \varphi_{DOA})$  (cf. Eq (6)).

The reference signal  $A_n^m(k)$  and reverb signal

$S_{n\text{Re}v}^m(k, r_a)$  will be ideally reproduced using the simple source approach [11]. The driving signal of each loudspeaker  $p$  for the reference signal is:

$$D_p(k) = -\frac{g_p(k)}{ikR^2} \sum_{n=0}^N \frac{e^{-ikR}}{h_n^{(2)}(kR)} \sum_{m=-n}^n A_n^m(k) Y_n^m(\theta_l, \varphi_l) \quad (11)$$

where  $g_p$  is a weighting term for loudspeaker number  $p$ .

In this ideal Ambisonics reproduction, it is assumed that the playback room has no reflections. In practice, the playback room will have reverberation that will deteriorate the quality of sound. In addition, there are always non-ideal frequency responses of loudspeakers. For this reason, an equalization must be performed to retain the quality and perceptual cues of a recording sound.

## 3. SIGNALS FOR OPTIMIZATION

### 3.1 Reference RIRs

A spherical microphones array with  $Q$  elements is used to capture and extract the Ambisonics coefficient of a room impulse response. The recorded SHs according to Eq (5) is called  $H_{n\text{rec}}^m(k, r_e)$ .

Using Eq (11), the driving signal of each loudspeaker by substitution of  $A_n^m(k) = H_{n\text{rec}}^m(k, r_e)$  is  $D_{ref}(k)$  and in time domain is denoted by  $d_{ref}[n] = F^{-1}\{D_{ref}(k)\}$ , where  $F^{-1}\{\}$  is the inverse Fourier transform operator.

The head-related-impulse-response (HRIR) of the left ear reconstructed by  $P$  loudspeakers in the anechoic room is:

$$BRIR_{ref,L}[n] = \sum_{p=1}^P d_{ref}[n] * hrtf_{p,L}[n]. \quad (12)$$

The derivation of the right ear signal is identical if  $L$  is replaced by  $R$  superscripts.  $BRIR_{ref,R}$  and  $BRIR_{ref,L}$  are used as the reference signals for optimization. From here on, the formulation is only written for the left ear.

### 3.2 Beamformed direct RinR impulse responses

The recorded beamformed RIR (BF) are convolved with BRIRs of the VBAP loudspeakers (SP1, SP2, SP3) to construct the binaural RinR impulse response for the beamformed direct sound ( $BRinR_{BF}$ ) for the left and

right ears of a listener in a reproduction room. Eq(5) is used for computation of  $H_{BF}(k, r_a)$ . When transformed to the time domain this signal is denoted by  $h_{BF}[n] = F^{-1}\{S_{BF}(k, r_a)\}$ . The beamformed signal reproduced by VBAP [6] received by the left ear is:

$$BRinR_{BF,L}[n] = h_{BF}[n] * \sum_{v=1}^3 g_v \times BRIR_{SP_v,L}[n], \quad (13)$$

where  $g_v$  is the VBAP gain of loudspeakers  $SP_v$ .

### 3.3 The reverb RinR impulse responses

In the previous section, the recording and reproduction of the direct sound component was discussed. For a perceptually accurate reproduction, also the reverberant sound field needs to be reproduced such that together with the direct sound that is reproduced the combined reproduced sound field will sound as similar as possible to original recorded sound field.

For this purpose, the remaining non-direct RIR of recording room in SHs domain is computed by subtraction of spatialized direct signal in DOA from original SHs coefficient using Eq (9). Only  $W(k) = S_{0\text{Re}v}^0(k, r_a)$  and  $Y(k) = S_1^{-1}\text{Re}v(k, r_a)$  that are the pressure and pressure derivative along y-axis, respectively are used as the loudspeaker  $p$  driving signal,  $1 \leq p \leq P$ . The y-axis is parallel to the connecting line between both ears of the listener. Using these two SHs coefficients, we have the following driving signals for each loudspeaker:

$$\begin{aligned} W_p(k) &= \frac{1}{ik} \frac{e^{-ikR} W(k)}{h_0^{(2)}(kR)} Y_0^0(\theta_l, \varphi_l) \\ Y_p(k) &= \frac{1}{ik} \frac{e^{-ikR} Y(k)}{h_1^{(2)}(kR)} Y_1^{-1}(\theta_l, \varphi_l) \end{aligned} \quad (14)$$

The equivalent time domain signals are

$$w_p[n] = F^{-1}\{W_p(k)\} \text{ and } y_p[n] = F^{-1}\{Y_p(k)\}.$$

The binaural RinR impulse response ( $BRinR$ ) of the reverb sound for the left ear denoted by  $BRinR_{\text{Re}v,L}$  is produced by summation of left ear RIRs convolved with  $w_p[n]$  and  $y_p[n]$  signals:

$$\begin{aligned} BRinR_{\text{Re}v,L}[n] &= \left( \sum_{p=1}^P w_p[n] * h_{p,L}[n] \right) + \left( \sum_{p=1}^P y_p[n] * h_{p,L}[n] \right) \\ &= BRinR_{\text{Re}v,W,L}[n] + BRinR_{\text{Re}v,Y,L}[n] \end{aligned} \quad (15)$$

## 4. OPTIMIZATION PROCEDURE

Using an optimization procedure, the spectral energy distributions of the direct and the reverb sound

components of the reproduced signal will be made equal to those of the reference signal. For the reference RIR of Eq (12), after separation of direct and reverb parts:

$$BRIR_{ref,L}[n] = BRIR_{ref,dir,L}[n] + BRIR_{ref,rev,L}[n], \quad (16)$$

and for the beamformed direct sound reproduced in the playback room of Eq (13):

$$BRinR_{BF,L}[n] = BRinR_{BF,dir,L}[n] + BRinR_{BF,rev,L}[n] \quad (17)$$

The direct and reverb parts are filtered by a 4th-order Gammatone filterbank described in [12]. The Auditory Transfer Function  $ATF_{sig}^{(i)}$  is defined as the energy of a signal at the output of  $i$ th filter of Gammatone filterbank. After equalization of direct signal, we have:

$$ATF_{ref,dir,left}^{(i)} = \left( g_{vbap,L}^{(i)} \right)^2 ATF_{BF,dir,L}^{(i)} \quad (18)$$

where  $g_{vbap,L}^{(i)}$  is the gain of VBAP loudspeaker in each frequency band. A method proposed in [13] is used for gain optimization in which the overlapping Gammatone filters are considered. This method also avoids negative gains. The average value of the left and right ear gains are considered as the final gain for the VBAP loudspeakers in each frequency band:

$$g_{vbap}^{(i)} = \frac{g_{vbap,L}^{(i)} + g_{vbap,R}^{(i)}}{2} \quad (19)$$

The reverberant RinR impulse response is including two parts of Eq (15) and the reverb part of Eq (17). Each term has its own weight:

$$\begin{aligned} BRinR_{Rev,L}^{(i)}[n] &= g_{vbap}^{(i)} BRinR_{BF,rev,L}^{(i)}[n] \\ &+ \alpha^i BRinR_{Rev,W,L}^{(i)}[n] + \beta^i BRinR_{Rev,Y,L}^{(i)}[n] \quad (20) \\ &= g_{vbap}^{(i)} BRinR_{BF,rev,L}^{(i)}[n] + BRinR_{Rev,AMB,L}^{(i)}[n] \end{aligned}$$

$g_{vbap}^{(i)}$  is the gain of beamformed direct sound that is also multiplied by reverb part of this signal. This is performed, because the direct and reverb parts of the VBAP loudspeakers of real signals are not separable. The  $\alpha^i$  and  $\beta^i$  are gains to control the IC between left and right ear signals in each frequency band. The IACC and the IC are defined as:

$$IACC^{(i)}[q] = \frac{\sum_{m=-\infty}^{m=\infty} BRIR_L^{(i)}[m] BRIR_R^{(i)}[m+q]}{\sqrt{\sum_{m=-\infty}^{m=\infty} \left( BRIR_L^{(i)}[m] \right)^2} \sqrt{\sum_{m=-\infty}^{m=\infty} \left( BRIR_R^{(i)}[m] \right)^2}} \quad (21)$$

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

The energy equalization for the reverb signal is separately performed using the  $g_{amb,L}^{(i)}$  gain in each frequency band:

$$\begin{aligned} ATF_{ref,rev,L}^{(i)} &= \\ \left( g_{vbap}^{(i)} \right)^2 ATF_{BF,rev,L}^{(i)} &+ \left( g_{amb,L}^{(i)} \right)^2 ATF_{Rev,AMB,L}^{(i)}. \end{aligned} \quad (22)$$

Similar to the gain averaging of VBAP loudspeakers, the average values of left and right ear gains are considered as the final gain of Ambisonic loudspeakers in each frequency band:

$$g_{amb}^{(i)} = \frac{g_{amb,L}^{(i)} + g_{amb,R}^{(i)}}{2} \quad (23)$$

The optimized final signal in each frequency band including direct and reverb sound is:

$$\begin{aligned} BRIR_{opt,L}^{(i)} &= \\ g_{vbap}^{(i)} BRIR_{ref,L}^{(i)}[n] &+ g_{amb}^{(i)} BRinR_{Rev,AMB,L}^{(i)}[n] \end{aligned} \quad (24)$$

The IC between  $BRIR_{opt,L}^{(i)}$  and  $BRIR_{opt,R}^{(i)}$  namely  $IC_{opt}^i$  is computed and compared with  $IC_{ref}^i$ . The error value in each frequency band is  $IC_{error}^{(i)} = |IC^{(i)} - IC_{ref}^{(i)}|$ . To find the best values for  $\alpha^i$  and  $\beta^i$ , a searching procedure repeatedly performs an optimization using Equations (20), (21), (22) and (23). The  $\alpha^i$  and  $\beta^i$  with a minimum value of the  $IC_{error}^{(i)}$  considering the derived  $g_{vbap}^{(i)}$  and  $g_{amb}^{(i)}$  are taken as the filter weights.

## 5. SIMULATIONS AND EVALUATION OF THE PROPOSED METHOD

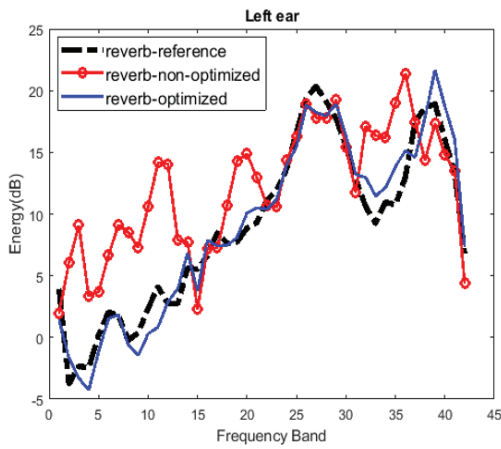
To simulate the recording room the SMIR toolbox [14] and for simulation of the playback room, the RAZR toolbox [15] are used. For both Ambisonics recordings and reproductions, a 50-node Lebedev Grid [16] and 4th order Ambisonics incorporating 25 SHs are simulated. The optimization is performed using 42 Gammatone filters. The parameters of two recording scenarios and a playback room are shown in Tables 1 and 2, respectively. The results of the energy equalization for reverb sound in scenario 2 for left and right ears are depicted in Figure 1 and Figure 2 respectively. Note that for the case that reverberation is not optimized, the reverberant signal extracted according to Eq (15) is simply played back without equalization such as would be done if the playback room were anechoic, leading to an increase of power in the low frequencies.

	room dim	src pos	mic pos	T60
Scenario 1	(50,20,6)	(6,15,1)	(16,15,1)	750ms
Scenario 2	(50,20,6)	(6,15,1)	(15,10,1)	750ms

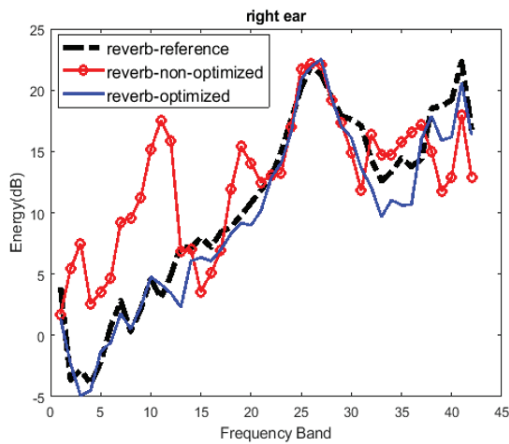
**Table 1.** Two recording scenarios. The room dimensions (dim), source position (src pos), microphone position (mic pos) and reverberation time (T60) are given.

	room dim	array center	array Radius	T60
playback	(8,5,5)	(2,2,2)	2 m	500ms

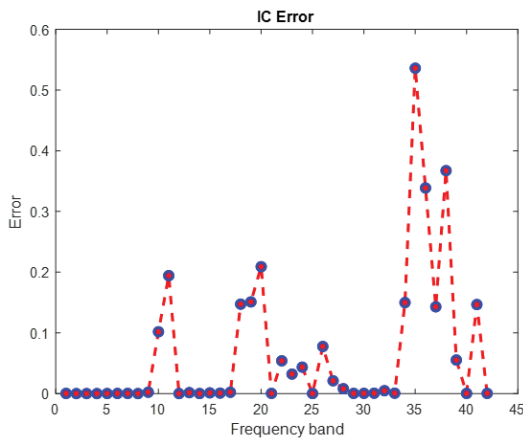
**Table 2.** Simulated playback room using a loudspeaker array with 50-nod Lebedev grid.



**Figure 2.** The ATF of reverb sound at the output of all Gammatone filters for the reference, non-optimized and optimized signals of left ear.



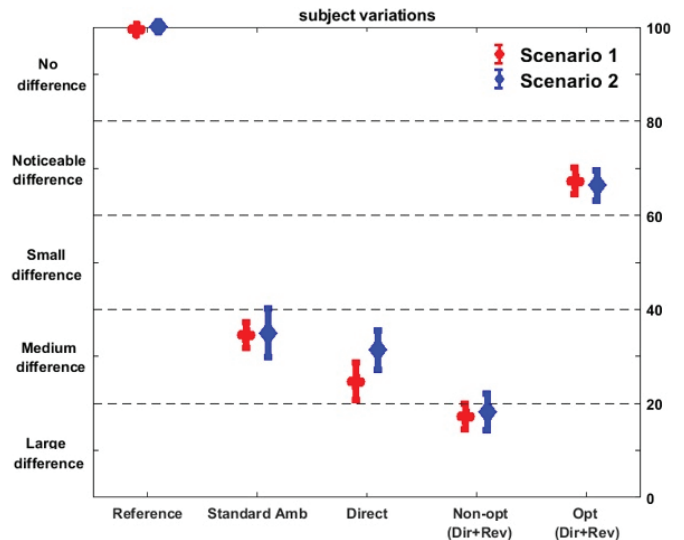
**Figure 2.** The ATF of reverb sound at the output of all Gammatone filters for reference, non-optimized and optimized signals of right ear.



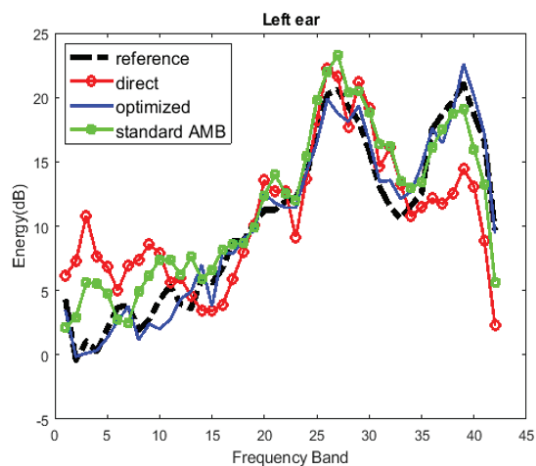
**Figure 3.** The IC Error after final compensation for scenario 2.

The IC errors after final compensation are depicted in Figure 3. It is seen that in some frequency bands, complete compensation is not possible and errors are inevitable.

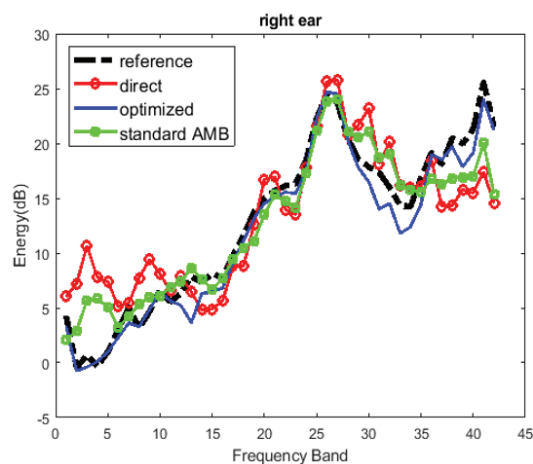
A multi-stimulus scaling experiment was performed with 8 listeners over headphones similar to the MUSHRA test [17]. The signals used in this experiment are an ideally reproduced reference signal with Ambisonics (Reference), a reproduced signal in the reverb room using standard higher-order Ambisonics without room optimization (Standard Amb), the beamformed direct signal played using VBAP method (Direct), the non-optimized signal consisting of direct signal produced by VBAP loudspeaker plus reverb signal produced by the loudspeaker array without any optimization for the playback room (non-opt (Dir+Rev)) and finally the proposed optimized signal (opt (Dir+Rev)). For each scenario, 4 different instrumental sounds and one human voice are played. The mean and standard errors of measurements are depicted in Figure 4. The standard Ambisonics reproduction including all 25 SHs in a reverb room (Standard Amb) leads to a high signal degradation compared to the reference signal. The non-optimized case (non-opt (Dir+Rev)) including direct sound and two of SHs ( $W(t)$  and  $Y(t)$  signals) shows the worst quality between all signals whereas, the our proposed optimization shows considerable enhancement. In the optimization method only two SHs coefficients are used which in some applications will substantially reduce the required data transmitted compared to a standard higher-order ambisonics transmission. Note also that the non-optimized beamformed direct sound (Direct) does not reproduce the desired spatial characteristics of the reference signal and its scoring is below the Standard Ambisonics.



**Figure 4.** Ratings of the multi stimulus scaling experiment over headphones. The circles and diamonds illustrate the results for scenario 1 and scenario 2, respectively. The mean-score and the standard error across the subjects are shown.



**Figure 5.** The ATF comparison between the reference, direct, optimized and standard Ambisonics signals of scenario 2 for left ear.



**Figure 6.** The ATF comparison between the reference, direct, optimized and standard Ambisonics signals of scenario 2 for right ear.

The ATF comparison between the direct, reference, optimized and standard Ambisonics signals for the left and right ears are shown in Figure 5 and Figure 6. The RMS values of direct and standard Ambisonics signals are equalized to that of the optimized signal. According to these figures, the playback room has smaller absorption coefficient for low frequencies for both direct and standard Ambisonic cases. After the optimization, these high-energy components in low frequencies are equalized. It is also shown that the direct signal, without optimization, in spite of incorporating the direction information only shows degraded quality, i.e. more low frequency energy, even in comparison with the standard Ambisonics signal.

## 6. SUMMARY

A new recording and reproduction method based on Ambisonics to perceptually compensate a playback room is proposed. Instead of ideally reconstructing the original sound field, the directional cues, energy and coherence of original sound are reproduced. The energy optimizations

are separately performed for direct and reverb sounds. The direct sound is recorded using beamforming and played back via the VBAP. For the reverb sound instead of recording and transmission of all Ambisonics coefficient, only the pressure (W signal) and one of derivatives of pressure (Y signal) are used and the optimization is performed at the playback side. Information of these two channels are sufficient to compensate for energy and IC of the reproduced signal to match with the recorded signal. The results of listening tests shows a strong improvement in comparison to the standard Ambisonics without compensation. A current limitation of this optimization method is the requirement of the impulse responses of the recording room. Another limitation is related to the handling of only one source in each time frame. Modifications are planned to extend this optimization framework for more than one concurrent sound source. In addition, the effect of sweet point displacement must be evaluated.

## ACKNOWLEDGEMENTS

We would like to thank the Deutsche Forschungsgemeinschaft for supporting this work as part of the Sonderforschungsbereich 1330 Hörakustik (HAPPAA)

## REFERENCES

- [1] A. J. Berkhout, "A holographic approach to acoustic control," *Journal of the audio engineering society*, vol. 36, no. 12, pp. 977-995, 1988.
- [2] M. A. Gerzon, "Periphony: With-height sound reproduction," *Journal of the audio engineering society*, vol. 21, no. 1, pp. 2-10, 1973.
- [3] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *The Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2100-2111, 2005.
- [4] P. Lecomte, P.-A. Gauthier, C. Langrenne, A. Berry, and A. Garcia, "Cancellation of room reflections over an extended area using Ambisonics," *The Journal of the Acoustical Society of America*, vol. 143, no. 2, pp. 811-828, 2018.
- [5] V. Pulkki, "Spatial sound reproduction with directional audio coding," *Journal of the Audio Engineering Society*, vol. 55, no. 6, pp. 503-516, 2007.
- [6] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the audio engineering society*, vol. 45, no. 6, pp. 456-466, 1997.
- [7] A. Politis, J. Vilkamo, and V. Pulkki, "Sector-based parametric sound field reproduction in the spherical harmonic domain," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 852-866, 2015.
- [8] J. Grosse and S. van de Par, "Perceptually accurate reproduction of recorded sound fields in a reverberant room using spatially distributed loudspeakers," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 867-880, 2015.
- [9] B. Rafaely, *Fundamentals of spherical array processing*. Springer, 2015.
- [10] A. Koretz and B. Rafaely, "Dolph-Chebyshev beampattern design for spherical arrays," *IEEE transactions on Signal processing*, vol. 57, no. 6, pp. 2417-2420, 2009.
- [11] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *Journal of the Audio Engineering Society*, vol. 53, no. 11, pp. 1004-1025, 2005.

- [12] V. Hohmann, "Frequency analysis and synthesis using a Gammatone filterbank," *Acta Acustica united with Acustica*, vol. 88, no. 3, pp. 433-442, 2002.
- [13] 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.
- [14] D. Jarrett, E. Habets, M. Thomas, and P. Naylor, "Rigid sphere room impulse response simulation: Algorithm and applications," *The Journal of the Acoustical Society of America*, vol. 132, no. 3, pp. 1462-1472, 2012.
- [15] 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.
- [16] P. Lecomte, P.-A. Gauthier, C. Langrenne, A. Berry, and A. Garcia, "A fifty-node Lebedev grid and its applications to ambisonics," *Journal of the Audio Engineering Society*, vol. 64, no. 11, pp. 868-881, 2016.
- [17] I. BS, "1534-1, Method for the subjective assessment of intermediate quality level of coding systems," *International Telecommunications Union, Geneva, Switzerland*, vol. 14, 2003.