DOI: 10.61782/fa.2023.0830

NEAR-FIELD SOURCE REPRODUCTION IN LARGE LOUDSPEAKER ARRAYS

Laurent S. R. Simon[1*] **Stefan Klockgether**[1]
[1] Sonova A.G., Stäfa, Switzerland

ABSTRACT

Spatial decomposition techniques can help to compensate for some of the limitations of higher order Ambisonics (HOA). Compared to HOA, higher order spatial impulse response rendering (HO-SIRR) has proven to substantially increase the sweet spot of a simulated space. For HO-SIRR-simulated scenes, it has also been shown that the performance of hearing aid beamformers is closer to the performance in real acoustic environments, and that interaural time and level differences are more similar to those recorded with an artificial head (KEMAR) in the original space than with a simulation based on simple HOA.

However, with spatial decomposition techniques, it is more difficult to simulate sources at various distances, especially when trying to simulate focused sources within a loudspeaker array.

In this presentation, we describe a near-field source rendering technique. HOA and hearing aid recordings were made in a reference room. The method presented in this paper was compared to HOA and HO-SIRR methods in a sound reproduction room, with regard to direct-to-reverberant ratio (DRR), coherence, and performance of the hearing aid beamformer.

Keywords: *spatial audio, higher order ambisonics,*

1. INTRODUCTION

The use of sound fields simulated with higher order Ambisonics (HOA) for the evaluation of hearing devices has shown a deterioration of interaural cues and the performance of hearing aid beamformers, compared to recordings made with hearing devices in the corresponding real sound fields. The differences are especially big when moving away from the center of the simulation [1, 2]. Simon et al. [3] showed that HOA impulse responses decoded using higher order spatial impulse response rendering (HO-SIRR) [4], significantly improved interaural cues and the performance of the beamformer compared to HOA sound reproduction. However, using HO-SIRR for sound reproduction means to not be able to simulate focused sources, i.e. sources emitting from inside the loudspeaker array [5]. When simulating a realistic environment in a large space, using HO-SIRR for target sources consequently means, that these target sources tend to appear to be further away from the subjects than desired. Additionally, the reverberation of the reproduction room might have a larger effect on the target sound characteristics. In such a case, it was hypothesized that the DRR and interaural cross-correlation (IACC) would be lower than intended.

In hearing research, recent studies aim to evaluate either the performance of hearing devices or the behaviour of hearing-impaired subjects in realistic conditions. Using realistic environment has a significant effect on the subject's speech reception thresholds [6]. More immersive environments also have effects on subject's behaviour, learning abilities and listening effort [7, 8]. Especially in studies where subjects are allowed to move to interact with acoustic scenes, it is crucial to optimize sound reproduction.

Several hybrid approaches were suggested in the past. Favrot and Buchholz (2010, [9]) used different orders of Ambisonics to reproduce different parts of a room impulse response. In the continuation of this, Weisser et al. (2019, [10]) provide a database of Ambisonics recordings and spatial impulse responses (SIRs). Each SIR is proposed either as a full impulse response or as

[*] ***Corresponding author***: *laurent.simon@sonova.com.*

two separated impulse responses containing only the direct part or only the rest of the impulse response. This was made to facilitate the hybrid decoding of the SIRs.

Pelzer et al. (2011, [11]) used a loudspeaker array of 8 loudspeakers to reproduce both the direct and early reflections with a crosstalk cancellation technique, and the diffuse sound via Ambisonics. However, the solution is applicable to only one listener at a time. Otto and Hamdan (2016, [12] proposed a similar idea, but designed the system to be able to play the same transaural signal to two users at the same time. Binaural synthesis, however, comes with a number of limitations, mainly caused by the need for individual HRTFs and the impossibility for subjects to use their own hearing-aids.

This article presents an alternative to the techniques mentioned above and to standard HOA and HO-SIRR for simulating near-field sources in a large sound reproduction space. It consists in adding loudspeakers inside the room and redirecting the direct part of the SIRs to these additional loudspeakers, while optimizing the HOA decoding of the rest of the impulse response using HO-SIRR. This new near-field source rendering (NFSR) method is compared to a reference recording, to HOA and to HO-SIRR reproductions, with regard to direct-to-reverberant ratio (DRR), interaural cross-correlation (IACC), and front-back ratio of a static monaural beamformer (FBR, the ratio between the level of the front sources and the level of the back sources).

2. DESCRIPTION OF THE ALGORITHM

In a large non-anechoic sound reproduction room, a loudspeaker setup composed of two ensembles is considered: a dense set of loudspeakers L_d, and an additional sparse set of loudspeakers L_s. The loudspeakers in L_d are located close to the borders of the room. Since the room is non-anechoic, an attempt to decode HOA room impulse responses using L_d could result in DRRs and IACCs lower than those of the original sources when recorded with an artificial head.

In this sound reproduction room, subjects are allowed to move around. Consequently, decoding the HOA impulse responses using a standard decoder [13] would result in significant interaural time differences (ITD) and interaural level differences (ILD) errors, caused by spatial aliasing. In addition to the direct consequence of altered ITDs and ILDs on human perception, the phase errors caused outside of the sweet spot by the HOA decoding can deteriorate the behaviour of features such as beamformers and localization algorithms. This limits the applicability of HOA in hearing device evaluation cases. Even when the user is positioned exactly at the sweet spot, the ears are already slightly off target. This leads to a drop of performance of algorithms above 1.5kHz [2].

For that reason, Simon et al. [3] considered using HO-SIRR to simulate acoustic scenes to limit the deterioration of cues. The performance of the beamformer, as well as the actual ITDs and ILDs, were still significantly lower compared to the reference when the reproduction was done for large rooms [3]. This was hypothesized to be caused by the reverberation of the reproduction room.

Additionally, one of the typical situations of interest in hearing research, is a subject trying to follow a conversation, such as in a cocktail party. Informal feedback on using HO-SIRR in a large space led to the conclusion that distance is not sufficiently well rendered. The subjects in that room reported to have the impression of following a conversation with people located a few meters away from each other.

In order to improve the issues mentioned above, a new algorithm was developed. The algorithm includes the sparse second set of loudspeakers L_s for sound reproduction, in addition to the main, dense set of loudspeakers L_d.

In order to decode an N^{th} order HOA room impulse response $s(t)$, the impulse response is first separated into two parts using HO-SIRR:

$$s(t) = s_{ndiff}(t) + s_{diff}(t) \tag{1}$$

s_{ndiff} contains the non-diffuse part of the impulse response, whereas s_{diff} is an HOA signal of the same order as s, containing the diffuse part of the room impulse response. Following the algorithm in McCormack et al. [4]. s_{diff} is decoded as an HOA signal using a sampling decoder. In order to improve diffuseness, the loudspeaker signals are subsequently decorrelated. The loudspeakers used for the decoding of s_{diff} are a subset of L_d. The subset is chosen to contain between $(N+1)^2$ and $(N+2)^2$ loudspeakers, as it is recommended to use more than $(N+1)^2$ to benefit from optimal localization while having as little loudspeakers as possible to reduce sound instabilities [14, 15].

The non-diffuse part of the impulse response s_{ndiff} is further separated as:

$$s_{ndiff}(t) = s_d(t) + s_r(t) \tag{2}$$

where s_d is the direct sound and s_r the non-diffuse room reflections. s_d and s_r are obtained using the algorithm described in Weisser et al. [10]. This algorithm separates the direct sound and the remaining of the impulse response using a frequency-dependant window size. s_r is decoded using Vector-Base Amplitude Panning, as described in [4], using the whole ensemble L_d to increase spatial precision of the room reflections. The direct sounds s_d are routed to the loudspeakers of the sparse ensemble L_s, which are closest to the intended direction of the source.

3. EVALUATION METHOD

The near-field source rendering was evaluated using a setup with KEF E301 co-axial loudspeakers surrounding a 5m-by-5m-by-2.7m reproduction space, which is included in a larger reproduction room. The dense set L_d is composed by 89 loudspeakers, which positions are shown in Fig. 1 (blue and orange dots). Their level and delay was compensated to ensure that all loudspeakers would have the same level and arrival time at the center of the system. The RT_{60} of the reproduction room, measured at the center of the reproduction space, was 160 ms. At the center of the reproduction space, the DRR for sounds coming from the loudspeakers in L_d varied between $-3.6\ dB$ and $12.4\ dB$. The distance of the loudspeakers to the center of the reproduction space varied depending on the positions in the setup. The levels and delays of all the loudspeakers were therefore compensated during the calibration process, to ensure that the arrival time and the level of an acoustic wave at the center of the reproduction space were independent of the loudspeakers used.

For the measurements, a KEMAR head and torso simulator was installed at the reference point, at the center of the room. Twelve additional KEF E301 loudspeakers used as additional ensemble L_s were positioned at 1.5 m distance from the KEMAR on ear height at every 30°. The KEMAR had anthropometric ears and was equipped with hearing aid shells that contained hearing aid microphones. The distance of 1.5 m was selected as the RIRs that were originally considered for that evaluation were recorded at the same distance. Although, for technical reasons, the RIRs actually used for that experiment were different.

The Room impulse responses (RIR) used as reference were measured in a reverberant environment (RT_{60} = 866 ms, DRR = 8.3 dB) using 5s-long exponential sweeps ranging from 100 Hz to 20 kHz [16]. In each of these rooms, 12 Genelec 8020 loudspeakers were placed every 30° at a height of 1.20 m and a distance of 2 m. The RIR were measured either using an Eigenmike EM32 microphone (leading to 4th order Ambisonics recordings) at the center of the array or a KEMAR wearing the hearing aid shells.

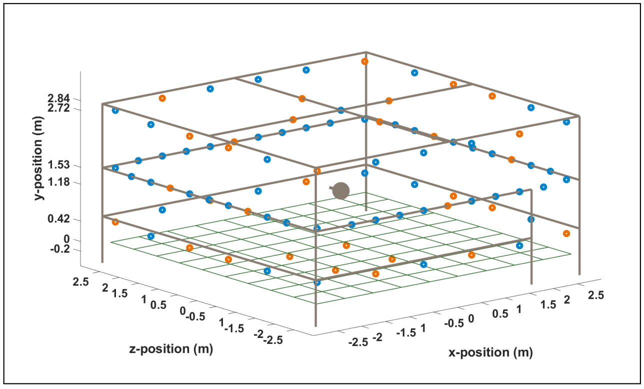

Figure 1. Position of the 89 loudspeakers of L_d in the setup. The orange dots indicate the positions of the 32 loudspeakers used for the HOA reproduction, whereas the HO-SIRR reproduction used the 89 loudspeakers marked by both the blue and orange dots.

The KEMAR recordings with hearing aid shells in the original space is used as a reference. The test conditions are described in Tab. 1, and were recorded using the same KEMAR and hearing aid shells.

Table 1. Test conditions

cond. name	description
reference	reference space
HOA	subset of L_d using 32 loudspeakers, allrad decoder
HO-SIRR	subset of L_d for s_{diff}, full L_d for s_{ndiff}
NFSR	subset of L_d for s_{diff}, L_s and full L_d for s_{ndiff}, as described above

For HOA decoding, Bertet (2009, [15]) showed that for practical applications with people, it is recommended to have a number of loudspeakers higher than $(M+1)^2$, where M is the HOA order. To optimize audio quality when users move their head, this number should not be much higher than $(M+1)^2$ [15]. Although KEMAR was not moved during recording, it was decided to decode the HOA stimuli over 32 loudspeakers. The position of the loudspeakers was calculated to be as equally distributed in the setup as possible, given by the positions of the 89 loudspeakers in the standard setup of the room. The HOA decoding was calculated in Matlab using the Higher Order Ambisonics toolbox by Archontis Politis [1] .

4. RESULTS

In a hearing device, it is often desirable to attenuate surrounding noise. One way to achieve this is through the use of a beamformer, which keeps the sounds from a target direction unmodified and attenuates the sounds coming from the other directions. The performance of the beamformer is measured as a Front-Back Ratio (FBR). For each direction of simulated room impulse response, the recordings were therefore analyzed in terms of direct-to-reverberant ratio (DRR), maximum of the interaural cross-correlation coefficient (IACC), and FBR. Results are shown in Fig. 2, Fig. 3, and Fig. 4 respectively.

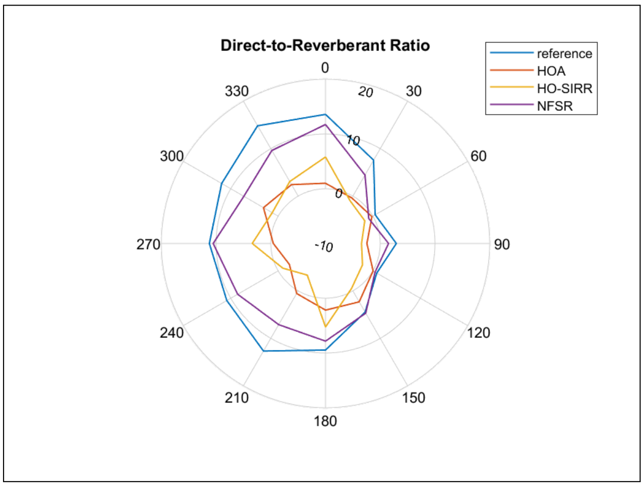

Figure 2. Direct-to-Reverberant Ratio, in dB, as a function of the direction of the sound, in the front left microphone.

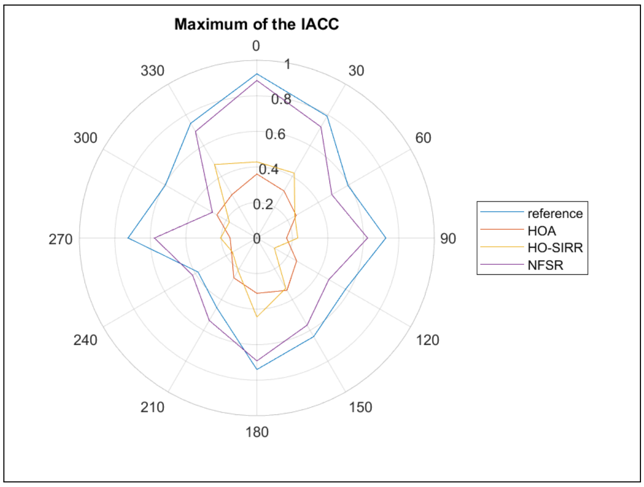

Figure 3. Maximum of the interaural cross-correlation, as a function of the direction of the sound, between the two front microphones.

The absolute differences between the reference DRR and the test system DRR on the front left microphone, averaged over all directions is 7.9 dB, 7.8 dB, and 2.4 dB for higher order Ambisonics (HOA), higher order spatial impulse response rendering (HO-SIRR), and the suggested near field source reproduction (NFSR) respectively (Fig. 2). Similarly, the absolute difference between the reference IACC and the test system IACC, calculated between the front microphones of the hearing aid shells and averaged over all directions, is 0.4, 0.37, and 0.09 for HOA, HO-SIRR, and NFSR respectively (Fig. 3). The absolute difference of FBR of the left hearing aid shell, averaged over all frequencies in the linear domain, is 3.1 dB, 2.8 dB, and 1.8 dB for HOA, HO-SIRR, and NFSR respectively (Fig. 4).

The NFSR shows a DRR, ICC and FBR performance closer to the reference, when compared to the HOA and HO-SIRR decoding in the non-anechoic room. This was expected, as the direct sound is emitted from a smaller distance when using NFSR. The evaluation was conducted only at the center of the room, and no formal perceptual evaluation has been conducted yet. However, the performance is expected to be better using NFSR as long as the distance to the near-fied loudspeaker emitting the sound is smaller than the distance to the surrounding loudspeakers emitting the sound. This should be the case in most possible, realistic positions of the subjects in the presented acoustic scenes.

[1] https://github.com/polarch/Higher-Order-Ambisonics

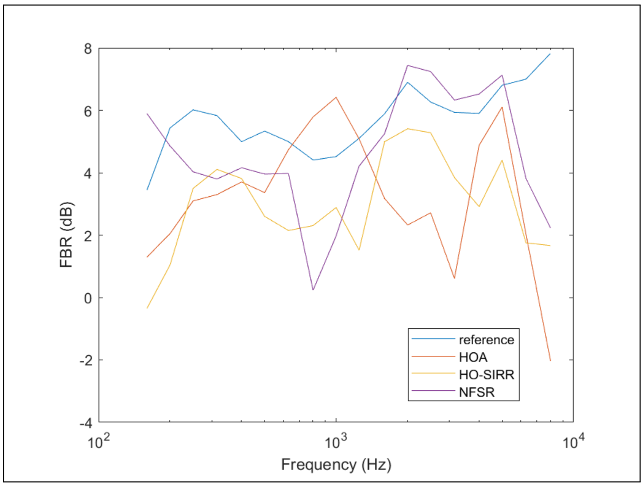

Figure 4. Front-back ratio for the reference, and the reproduced sounds via HOA, HO-SIRR, and NFSR, in the left beamformer output.

NFSR requires the addition of more loudspeakers inside a room, thus making movement potentially more complicated. An important constraint is that the additional loudspeakers should be placed at the position of the desired additional near field sound source. When moving away from the center, the relative direction of the early reflections in the simulated room compared to the relative position of the near field sound source will not match the relative positions in the original space.

5. CONCLUSION

In this article, a new method was proposed for simulating near-field sources in a large sound reproduction system. Measures using a KEMAR manikin showed that the new method leads to simulated sources more similar to the reference than sources decoded via HOA or HO-SIRR, in terms of DRR, IACC, and FBR. Further evaluation should compare different simulated rooms at the sweet spot and away from the sweet spot.

6. REFERENCES

[1] G. Grimm, S. Ewert, and V. Hohmann, “Evaluation of Spatial Audio Reproduction Schemes for Application in Hearing Aid Research,” *Acta Acustica united with Acustica*, vol. 101, pp. 842–854, July 2015.

[2] L. Simon, N. Dillier, and H. Wüthrich, “Comparison of 3D Audio Reproduction Methods Using Hearing Devices,” *Journal of the Audio Engineering Society*, vol. 68, no. 12, pp. 899–909, 2020.

[3] L. Simon, H. Wüthrich, and N. Dillier, “Investigating Higher Order Spatial Impulse Response Rendering for the evaluation of hearing devices,” in *DAGA*, (Wien), Aug. 2021.

[4] L. McCormack, V. Pulkki, A. Politis, O. Scheuregger, and M. Marschall, “Higher-Order Spatial Impulse Response Rendering: Investigating the Perceived Effects of Spherical Order, Dedicated Diffuse Rendering, and Frequency Resolution,” *Journal of the Audio Engineering Society*, vol. 68, pp. 338–354, June 2020.

[5] J. Ahrens and S. Spors, “Focusing of Virtual Sound Sources in Higher Order Ambisonics,” in *Proc. of the 124th Audio Engineering Society convention*, Audio Engineering Society, May 2008.

[6] N. Mansour, M. Marschall, A. Westermann, T. May, and T. Dau, “Speech Intelligibility in a Realistic Virtual Sound Environment,” *Proceedings of the 23rd International Congress on Acoustics*, pp. 7658–7665, 2019.

[7] L. Picinali, G. Grimm, Y. Hioka, G. Kearney, D. Johnston, C. Jin, L. S. R. Simon, H. Wüthrich, M. Mihocic, P. Majdak, and D. Vickers, “VR/AR and hearing research: current examples and future challenges,” in *Proc. of the Forum Acusticum 2023*, (Torino, Italy), Sept. 2023.

[8] J. Seitz, K. Loh, S. Nolden, and J. Fels, *Investigating Intentional Switching of Spatial Auditory Selective Attention in an Experiment with Preschool Children*. Mar. 2023.

[9] S. Favrot and J. M. Buchholz, “LoRA: A Loudspeaker-Based Room Auralization System,” *Acta Acustica united with Acustica*, vol. 96, pp. 364–375, Mar. 2010.

[10] A. Weisser, J. M. Buchholz, C. Oreinos, J. Badajoz-Davila, J. Galloway, T. Beechey, and G. Keidser, “The Ambisonic Recordings of Typical Environments (ARTE) Database,” *Acta Acustica united with Acustica*, vol. 105, pp. 695–713, July 2019.

[11] S. Pelzer, B. Sanches Masiero, and M. Vorländer, “3D reproduction of room acoustics using a hybrid syste...

- RWTH AACHEN UNIVERSITY Institute for Hearing Technology and Acoustics - English," in *Proceedings of ICSA 2011*, Nov. 2011.

[12] P. Otto and E. Hamdan, "Bridging Near and Far Acoustical Fields: a Hybrid Systems Approach to Improved Dimensionality in Multi-Listener Spaces," Audio Engineering Society, July 2016.

[13] F. Zotter, H. Pomberger, and M. Noisternig, "Ambisonic Decoding with and without Mode-Matching: A Case Study Using the Hemisphere," in *2nd Int. Symposium on Ambisonics and Spherical Acoustics*, (Paris, France), pp. –, May 2010.

[14] R. Nicol, "Sound Spatialization by higher order Ambisonics: encoding and decoding a sound scene in practice from a theoretical point of view," in *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, (Paris), May 2010.

[15] S. Bertet, *Formats Audio 3D Hiérarchiques : Caractérisation Objective et Perceptive Des Systèmes Ambisonics D'Ordres Supérieurs*. PhD thesis, Institut national des Sciences Appliquées de Lyon, Jan. 2009.

[16] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique," in *Audio Engineering Society 108th convention*, Feb. 2000. Preprint 5093.