

# Perceptual compensation of Eigenmike recordings to reproduce a spatial sound-field in a reverberant room

Ali Fallah<sup>1</sup>, Steven van de Par<sup>1</sup>

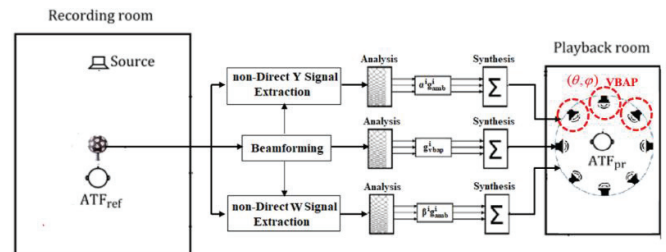
<sup>1</sup>Acoustics group, CvO University Oldenburg, Germany, Cluster of Excellence – Hearing4all,  
E-mail: ali.fallah@uni-oldenburg.de

## Introduction

In some applications, it is desirable to reproduce a recorded sound, for example in a concert hall, in another room in a way that the listeners in the playback room perceive the sound field in the same way as within the concert hall. Methods such as wave field synthesis (WFS) [1] and Ambisonics [2], in theory, would allow to perfectly reproduce a recorded sound field. But in practice they can have some limitations. The reconstruction area, specifically for high frequencies, is limited to a small area called sweet spot. Also, a need for large number of loudspeakers makes it difficult to use these methods as a commercial option. Moreover, it is assumed that the playback room is ideally anechoic. Beside these physically-based approaches for rendering and reproduction, perceptually motivated approaches such as directional audio coding (DirAC) [3] are proposed which do not necessarily reproduce the physical sound field, but allow for a perceptually accurate reproduction nevertheless. In the DirAC method, the direct and reverb parts of a sound field are recorded and processed separately and are played back using VBAP [4] and using a decorrelation method. Here, the effect of reverberation of the playback room is again neglected. Altogether, the effect of reverberation of the playback room is given relatively little attention in the literature. In this study, a perceptually-motivated method [5] for compensation of a reverb playback room is presented and evaluated in a listening test. The spectral envelope of direct and reverb sounds and also the interaural coherence (IC) of the original recorded sound are preserved by using a perceptually-based optimization. For this purpose, sound fields are compared and optimized between a dummy head in the recording room and another dummy head placed in the playback room at the listener positions.

## Method

The structure of the proposed Ambisonic-based recording and reproduction approach is shown in Figure 1. In the recording room, an Eigenmike microphone array [6] and in the reproduction room, a loudspeaker array is used. Because of the reverberation of the playback room, it is difficult to control the sound field. Therefore, similar to the DirAC, the direct and reverb sounds are recorded and compensated separately before playback. The perceptual parameters used for optimization are the spectral energies of the direct and reverberant parts and the interaural coherence (IC). These optimization parameters were also previously used in [7]. The binaural room impulse responses (BRIRs) of both recording and reproduction rooms are required for optimization. The direct sound is recorded using beamforming and after compensation, this signal, is played back using VBAP using information about to direction of arrival (DOA) of the direct

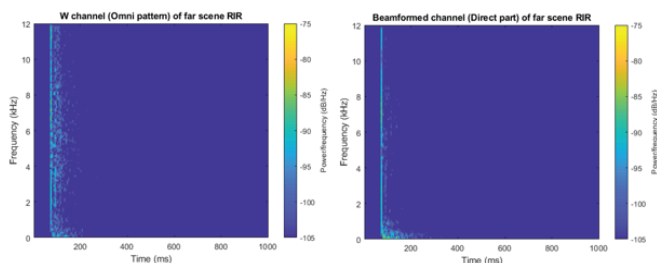


**Figure 1:** The block diagram of the room compensation method [5]. The signal is recorded using an Eigenmike and after separation to the direct and reverb parts and compensation using Gammatone analysis and synthesis framework are reproduced in a playback room. The direct sound is recorded by beamforming and after compensation is reproduced using VBAP. Two Ambisonic channels (W and Y) are used to record and compensate the reverb sound and are played back using conventional Ambisonics. BRIRs measured on two dummy heads in the recording and playback rooms are used as the reference and target signals for compensation.

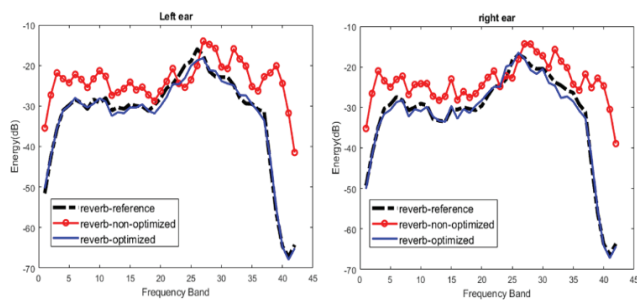
sound. The compensation of the spectral envelope is performed using a Gammatone analysis and synthesis framework [8] which allows to adjust the ‘Auditory’ Transfer Function measured at the dummy heads in the Recording and Playback room to be equal. The gain of Gammatone filter  $g_{vbap}^i$  is used for the spectral energy equalization of direct sound within filter number  $i$ . It is to be expected that this equalizing effectively compensates for effects of the microphones and loudspeakers, but not for the room. For recording and optimization of the reverb sound, the reverberation of the recording and playback rooms needs to be taken into account. For the reproduction of the diffuse sound field only two Ambisonic channels are used; the zero-order Ambisonics (W channel) and one first-order Ambisonic channels, with a dipole pattern (Y channel). The Y channel data is used to create an out-of-phase signal in left and right ears of the dummy head that is mixed with the W channel to obtain the desired interaural coherence. This is done by using a weighted summation of W and Y channels, and the weights of these two channels are shown by  $\alpha^i$  and  $\beta^i$ , respectively. Energy compensation for the reverb sound is performed by an additional weighting  $g_{amb}^i$  for the W and Y channels. For playback of the compensated reverb sound, all the loudspeakers are used. Note that the direct-sound field component is subtracted from the W and Y channels.

### Simulations and Measurements

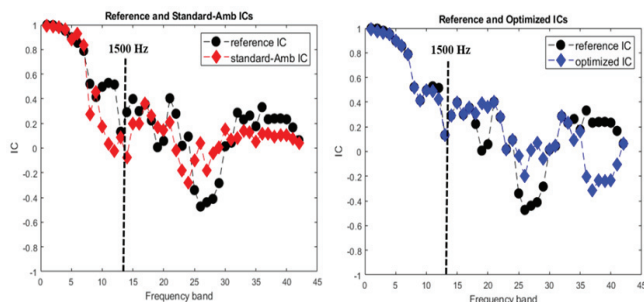
For evaluating the proposed method, Eigenmike recordings were made in a lecture room at different distances (3 and 8 meters) between the loudspeaker and the microphone array which are named the “close scene” and the “far scenes”. The playback room is simulated using the RAZR toolbox [9]. The 50 nodes Lebedev Grid [10] is used for positioning the loudspeakers in the simulated playback room. The reverberation time of both recording and simulated playback room is about 1 second. To record the direct sound and reduce the sidelobe level, an axis symmetric beamformer with Dolph-Chebyshev weightings [11] is used. For the Eigenmike recording of direct sound, a practical limitation of beamformer must be considered. Higher-order Ambisonics signals are restricted to increasingly limited operating frequency range in order to limit the maximum amount of the system self-noise generated during the encoding process [6]. For the lower frequencies, specifically frequencies below 400 Hz, the beampattern used to record the direct sound is therefore wider and as a consequence there is a higher amount of the reverb sound energy leaking to the direct-sound signal which may limit the ability of the proposed method to adjust the effective direct-to-reverberant energy to the desired value in the playback room. To evaluate the technical performance of the algorithm, results are shown that relate to the far scene. In Figure 2, the outputs of the omni-directional W channel and also the beamformer output are shown. It can be seen that the beamformer is able to suppresses the reverb sound well across a large range of frequencies. Only at low frequencies, because of the wider pattern of the beamformer, there is energy leakage of the reverb signal into the beamformer. This energy, to some extent can be accommodated for by the compensation procedure which will simply reduce the amount of reverberant energy ‘injected’ in the playback room by the W and Y channels. In Figure 3, the importance of compensation for the reverb sound in the playback room is shown. The reverb sound energies of left and right ears of dummy head at the output of each Gammatone after compensation are plotted and compared with that of the reference, non-optimized signals. In Figure 4, the ICs of the conventional (standard) Ambisonics and Optimized signal in comparison to that of the reference signal are depicted. Perceptually important frequencies in IC are at frequencies below 1500 Hz where the auditory system is most sensitive to changes in IC.



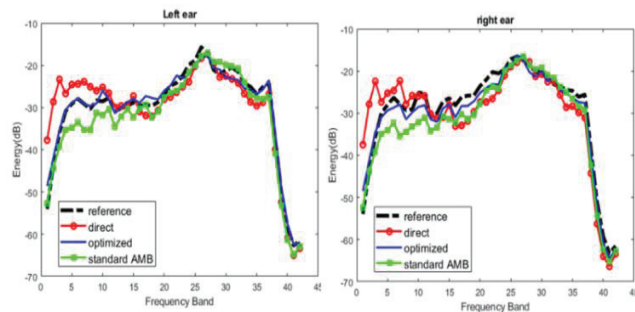
**Figure 2:** The outputs of W channel (left) and beamformer (right). In the high frequencies the beamformer suppresses the reverb sound well, but in the frequencies below 400 Hz because of wider pattern of the beamformer, there is energy leakage of the reverb signal into the beamformer.



**Figure 3:** The reverb sound energies of left and right ears of dummy head at the output of each Gammatone filter for the reference, non-optimized and optimized signals. It is clear that the optimized reverb signal is well fitted to the reference.

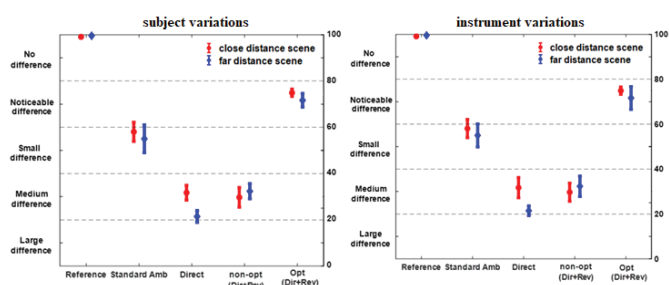


**Figure 4:** The results of IC compensation for perceptually important frequencies below 1500 Hz. After compensation (left) there is better matching between ICs of reference ICs and that of the compensated signal in comparison to the conventional Ambisonics (right).



**Figure 5:** The overall normalized energies of the reference, conventional Ambisonics, direct and optimized signals for the left and right ears. The optimized signal in comparison to the direct and the conventional Ambisonics is more fitted to the energy curve of the reference signal specifically at the low frequencies that the reverberation is dominant.

It is seen that all of Gammatone filter bank outputs in this range show a near to ideal compensation of ICs, much better than the IC found for conventional Ambisonics without room compensation. In the Figure 5, the normalized overall energies of the reference, conventional Ambisonics, direct and optimized signals for the left and right ears of the simulated dummy head are shown. The optimized signal is closer to the energy curve of the reference signal, specifically at the low frequencies it can be seen that the reverberation is dominant in comparison to the direct signal and the conventional Ambisonics. To also perceptually evaluate the proposed method, a MUSHRA-like listening test [12] was performed. In the listening test, 10 listeners participated and various types of audio signals were played. In Figure 6, the



**Figure 6:** The mean and standard deviation of the measurements considering subject variations (left) and instrument variations (right). The hidden reference signals are scored approximately 100% for all of listeners. Because of the reverberation of the playback room, the conventional Ambisonics is not successful and there is degradation in the similarity with the reference. The direct sound and the non-optimized signals have also lower similarity scores. The compensated signal that is only using 3 channels is superior than the conventional Ambisonics with 32 channels.

mean and standard deviation of the measurements are shown. In this MUSHRA test, the sound field recorded on the dummy head in the recording room served as a reference. It was compared against simply rendering the Ambisonics signal directly in the playback room (Standard Amb), by playing only the direct sound field component (resulting from the beamformer) in the playback room (Direct), by playing the direct and reverberant signal components in the playback room, without any compensation (non-opt Dir+Rev), and by playing the direct and reverberant signal components with the proposed optimization (Opt Dir+Rev). As can be seen, the hidden reference signals are well detected by the listeners and scored 100% almost for all of listeners. Because of the reverberation of the playback room, the Standard Ambisonics was judged to have more than noticeable differences compared to the reference signal. The direct dry sound (Direct) showed poorest performance and is clearly not enough to create a good reproduction. Reproducing both direct and reverberant components without optimization (non-opt, Dir+Rev) did not significantly improve quality. But after the proposed optimization (Opt. Dir+Rev), these signals were rated superior to the conventional Ambisonics without room compensation. These results show the importance of compensation in a reverb room. Interesting to point out is that the proposed method requires only 3 signals; the beamformed, direct sound signal, and the W and Y spherical harmonic components.

## Summary and Future Work

In this study, real Room-Impulse-Response recordings of an Eigenmike microphone array were used for evaluation of the perceptually motivated Ambisonic-based recording and reproduction approach. The room-impulse-response (RIR) in the presence of a sound source is recorded with a spherical microphone array. After obtaining the Eigenbeams and beamformed-RIR, direct and reverb parts of RIR are separated. Based on the recorded direct and reverb RIRs and the knowledge of BRIRs of the playback room, an optimization is performed resulting in a set of filter parameters. These are used in an analysis and synthesis

framework which includes filtering to compensate the direct and reverb sounds separately and which optimizes the spatial diffuseness. Quality assessments show that by using this compensation approach, the spectral and spatial properties of the original recorded sound are improved in comparison to conventional Ambisonics.

Our Next step is evaluation of robustness of the proposed approach to changing listening positions. Also, until now only one source is used in the recording scenario. Increasing the number of sources, extraction of direct and reverb sounds of each source, compensation of each source and finally the reproduction is the next challenge.

## Acknowledgements

We would like to thank the Deutsche Forschungsgemeinschaft for supporting this work as part of Project C2 of the Sonderforschungsbereich 1330 Hörakustik (HAPPAA)

## References

- [1] Berkhout, A. J.: A holographic approach to acoustic control, *Journal of the audio engineering society* (1988), vol. 36, no. 12, pp. 977-995.
- [2] Gerzon, M. A.: "Periphony: With-height sound reproduction," *Journal of the audio engineering society* (1973), vol. 21, no. 1, pp. 2-10.
- [3] Pulkki, V.: Spatial sound reproduction with directional audio coding, *Journal of the Audio Engineering Society* (2007), vol. 55, no. 6, pp. 503-516.
- [4] Pulkki, V.: Virtual sound source positioning using vector base amplitude panning, *Journal of the audio engineering society* (1997), vol. 45, no. 6, pp. 456-466.
- [5] Fallah, A. and van de Par, S.: A new approach for rendering Ambisonics recordings on loudspeakers in a reverberant environment (2020), *Forum Acusticum*, pp. 2071-2077.
- [6] Em32 Eigenmike Microphone array data sheet, mh acoustics (2013).
- [7] Grosse, J. and van de Par, S.: Perceptually accurate reproduction of recorded sound fields in a reverberant room using spatially distributed loudspeakers, *IEEE Journal of Selected Topics in Signal Processing* (2015), vol. 9, no. 5, pp. 867- 880.
- [8] Hohmann, V.: Frequency analysis and synthesis using a Gammatone filterbank, *Acta Acustica united with Acustica* (2002), vol. 88, no. 3, pp. 433-442.
- [9] Wendt, T., van de Par, S., and Ewert, S. D.: A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation, *Journal of the Audio Engineering Society* (2014), vol. 62, no. 11, pp. 748-766.
- [10] Lecomte, P., Gauthier, P. A., Langrenne, C., Berry, A. and Garcia, A.: A fifty-node Lebedev grid and its

- applications to ambisonics, *Journal of the Audio Engineering Society* (2016), vol. 64, no. 11, pp. 868-881.
- [11] Koretz, A. and Rafaely, B.: Dolph–Chebyshev beam pattern design for spherical arrays, *IEEE transactions on Signal processing* (2009), vol. 57, no. 6, pp. 2417-2420.
- [12] I. BS, "1534-1: Method for the subjective assessment of intermediate quality level of coding systems, International Telecommunications Union, Geneva, Switzerland (2003), vol. 14.