Exploring the Impact of Varying Number of Ambisonic Channels on a Perceptually-Based Audio Reproduction in Reverberant Rooms

Ali Fallah[1],Shun Nakamura, Steven van de Par[1]

[1]*Acoustics group, CvO University Oldenburg, Germany, Cluster of Excellence – Hearing4all,*

E-mail: *ali.fallah@uni-oldenburg.de*

Introduction

Ambisonics in audio technology has emerged as a prominent technique for capturing and rendering complex sound fields, promising an immersive reproduction of the original recorded audio in free-field playback scenarios [1]. In a conventional Ambisonics, it is assumed that the playback room is ideally anechoic. Nevertheless, the conventional application of this method are generally insufficient due to the order limitation and distorting influence of the playback room's acoustics, resulting in a perceptible decline in sound quality. Beside these physically-based approaches for rendering and reproduction, perceptually motivated approaches such as directional audio coding (DirAC) [2] are proposed which do not necessarily reproduce the physical sound field, but allow for a perceptually accurate reproduction nevertheless. In this study deals with this challenge by employing a perceptually-based optimization approach [3]. The energy-compensated direct and reverberant sounds with preserving the interaural coherence (IC) of the original recorded audio are rendered over 50 loudspeakers in a reverb room. For rendering of the reverb part, an effect of using varying number of first-order Ambisonics channels (2, 3 or 4 channels) is also investigated. The results of a quality-assessment experiment show a significant enhancement in the timbral and spatial aspects of the rendered audio for the compensated signals in comparison to the conventional Ambisonics. For the evaluation, signals of a dummy head in the recording room are compared with that of another dummy head placed in the playback room at the listener positions.

Method

The structure of the proposed Ambisonics-based recording and reproduction approach is shown in Fig 1. In the recording room, an Eigenmike microphone array [4] 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 [3,5]. 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 [6] using information about to direction of arrival (DOA) of the direct sound. The compensation of the

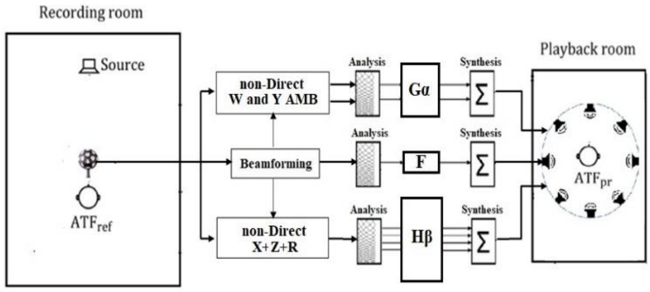

Figure 1: The block diagram of the room compensation method . The signal is recorded using an Eigenmike and after separation to the direct and reverb parts and compensation using Gammatone analysis and synthesis framework are reproduced in a playback room. The direct sound is recorded by beamforming and after energy compensation is reproduced using VBAP. Two Ambisonics channels (W and Y) plus other optional Ambisonics channels (X, Y, R) 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.

spectral envelope is performed using a Gammatone analysis and synthesis framework [7] which allows to adjust the Auditory Transfer Function (ATF) measured at the dummy heads in the recording and playback rooms to be equal. The gain of Gammatone filters in matrix **F** is used for the spectral energy equalization of direct sound within all Gammatone filters. 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, we have used previously only two Ambisonics channels; the zero-order Ambisonics (W channel) and one first-order Ambisonics channels, with a dipole pattern (Y channel) [3]. 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 mixing gains of these two channels are shown by matrix **G**. Energy compensation for the reverb sound is performed by an additional weighting $\boldsymbol{\alpha}$ for the W and channels. In this study we used extra dipoles (X and Z) and also all other ramming Ambisonics channels (R) for the rendering of reverb sound. The gain matrix for the mixing and energy are depicted by **H** and $\boldsymbol{\beta}$ respectively. For playback of the compensated reverb sound, all of loudspeakers are used.

Evaluation

For evaluating of the proposed method, Eigenmike recordings in a theater and a church are used. For the reproduction, a loudspeaker array using 50 nodes Lebedev Grid [8] is used for rendering. The reverberation time of theater and church are 0.5 and 1.2 second respectively. The reverberation time of playback room is in about 0.5 second. A MUSHRA-like listening test [9] with 13 participants for perceptually evaluation of the proposed method was performed. Various types of audio signals including snare drum, guitar, clarinet, piano, violin and speech are used.

In Fig 2, the 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 the Anker (low-passed version of the reference signal), the rendering using conventional Ambisonics signal directly in the playback room (CoAmb), an ideally reproduction of Eigenmike recordings in simulations (IdAmb), by playing only the compensated direct sound field component (resulting from the beamformer) in the playback room (VBAP), and by playing the direct and reverberant signal components with the proposed optimization using different number of Ambisonics channels (WY, WYX, WYXZ, WYXZR and WYZ). The WYXZ case is called B-format or first-order Ambisonics and the WYXZR case shows 4-order Ambisonics in which, the abbreviation ”R” stands for the remaining channels. As can be seen, the hidden reference signals are detected by the listeners and scored near to 100% for all of listeners. Because of the reverberation of the playback room, the conventional Ambisonics (CoAmb) was judged to have very high

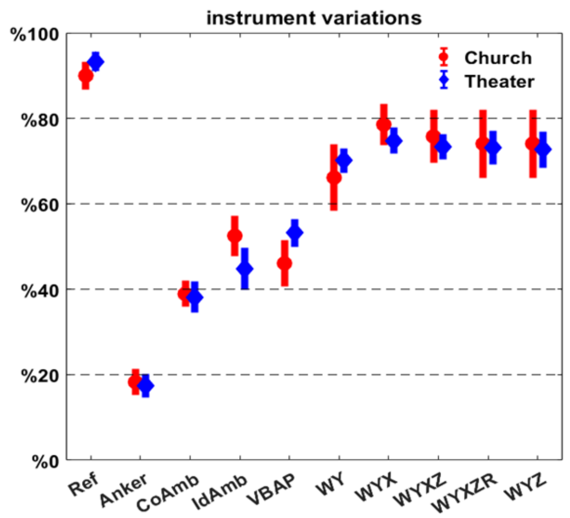

Figure 2: The mean and standard deviation of the measurements considering instrument. The hidden reference signals are scored in average near to 100% for all of listeners. Because of the reverberation of the playback room, the conventional Ambisonics (CoAmb) is not successful and there is degradation in the similarity with the reference. The only direct sound (VBAP) and interestingly also the Ideal Ambisonics (IdAmb) case have also lower similarity scores. The compensated signal using different number of channels (WY, WYX, WYXZ, WYXZR and WYZ) are superior than the other reproduced signals.

differences compared to the reference signal. Interestingly, the ideal Ambisonics reproduction show also very low quality in comparison to the reference signal. This is mainly because of limitation in reproduction of high frequencies in low-order Ambisonics that in our case the Ambisonics order is equal to four. The direct dry sound (VBAP) shows better performance in comparison to the conventional Ambisonics (CoAmb), but still is clearly not enough to create a good reproduction. Reproducing both direct and reverberant components with optimization using only WY components significantly improve quality. But after adding one additional Ambisonics channels (WYX and WYZ), first-order Ambisonics channels (WZXZ) and 4-order Ambisonics channels (WYXZR), significant improvements are seen in comparison to the WY case. The WYX case shows the highest mean score between the compensated cases. The results show the importance of compensation in a reverb room and using at least three Ambisonics channels for the reproduction.

Summary and Future Work

In this study, the effect of using different number of Ambisonics channels in the reproduction of reverb sound field was investigated. For this purpose, the real recordings of an Eigenmike microphone array were used for evaluation of the perceptually motivated Ambisonics-based recording and reproduction. Based on the recorded direct and reverb RIRs and the knowledge of BRIRs of the playback room, an optimization is performed to enhance the energy and interaural coherence of direct and reverb parts resulting in a set of Gammatone filter gains. We have previously shown that using of this optimization using only W and Y channels can improve the quality in comparisons to the uncompensated case using all 32 channels [3]. In this study, the quality assessments showed that by using at least three Ambisonics channels (WYX and WYZ), the spectral and spatial properties of the original recorded sound are improved in comparison to WY case. No significant improvement using first order Ambisonics (WYZX) and 4-order Ambisonics (WYXZR) in comparison to WYX and WYZ cases were observed. Our next step is an evaluation of robustness of the proposed approach using real-loudspeaker listening in the middle of array.

Acknowledgements

This work was supported by the DFG Cluster of Excellence (ID: 352015383–SFB 1330 C2, HAPPAA)

References

[1] Gerzon, M. A.: Periphony: With-height sound reproduction," Journal of the audio engineering society (1973), vol. 21, no. 1, pp. 2-10.

[2] Pulkki, V.: Spatial sound reproduction with directional audio coding, Journal of the Audio Engineering Society (2007), vol. 55, no. 6, pp. 503-516.

[3] 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.

[4] Em32 Eigenmike Microphone array data sheet, mh acoustics (2013).

[5] 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.

[6] 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.

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

[8] 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.

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