Spatial Resolution of Diffuse Reverberation in Binaural Ambisonic Playback

Djordje Perinovic[1] and Matthias Frank[2]

[1] University of Music and Performing Arts Graz, Austria, Email: djordje.perinovic@student.kug.ac.at

[2] Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Austria, Email: frank@iem.at

Introduction

Virtual acoustics for a single listener typically uses headphones, especially in virtual reality (VR) applications in combination with video headsets. Thus, the aim is the creation of binaural signals that are either authentic or plausible [1]. The simplest way of creating binaural signals is by measuring binaural room impulse responses (BRIRs) with a dummy head. In order to facilitate listener movement in the virtual scene, the measurements have to be done for multiple head orientations and positions [2]. Apart from the enormous measurement effort, the amount of data can be a problem for VR applications with limited resources, e.g. on mobile devices. The more sparse the measurement grid, the more severe becomes the challenges for interpolation [3, 4]. Interpolation for the incorporation of head rotation is much easier when using Ambisonics [5]. This is why Ambisonics is typically used in applications with single perspective and variable orientation, such as 360° videos. Another advantage of Ambisonics is the scalability of the spatial resolution in terms of the maximum order to adjust the playback to available computational resources.

The requirements for spatial resolution seem to be depending on the application and virtual environment: For a plausible playback of sound sources in a virtual room, 3^{rd} order might be enough [6], while authentic playback under anechoic conditions requires much higher orders [7]. Through psychoacoustically-motivated optimization, state-of-the-art binaural Ambisonic decoders [8, 9] can help to reduce the required order. Moreover, studies report that authentic playback of sources in reverberant rooms can be achieved with lower orders [10, 11]. These are interesting results, especially as the longer impulse responses of reverberant rooms pose a challenge for efficient convolution. It seems reasonable that the required spatial resolution reduces towards later parts of impulse responses.

In order to examine this hypothesis, this contribution evaluates the required spatial resolution under completely diffuse conditions, as they can be assumed after the mixing time of the spatial room impulse response [12]. The evaluation compares 0^{th}- to 6^{th}-order playback to a 7^{th}-order reference in a listening experiment. The experiment uses speech and a transient drum sound as audio signals and the Magnitude Least Squares (MagLS) approach [8, 9] for decoding to headphones. The diffuse reverberation is generated by a feedback delay network (FDN) [13, 14, 15].

The paper is structured as follows: First, the experimental setup and conditions are introduced, followed by the results and their interpretation. Finally, the paper is summarized and an outlook to further research is presented.

Setup and Conditions

The diffuse reverberation was created using the FdnReverb plug-in[1]. The reverberation time was set to 1.2 s and 6.7 s, respectively. For both reverberation times, the common settings were as follows: Room Size 20, Dry/Wet 1.0, Fdn Size 64, Fade In 0.21 s, and a high-shelf filter at 2 kHz with Q = 2 and -10 dB/s to reduce the reverberation time towards high frequencies (approx. 0.9 s and 3 s, respectively). In order to avoid a loss of diffusivity for lower order playback due to the reduced number of channels [16], all 64 channels of the output have been encoded into Ambisonics for all orders. The encoding matrices for the 64 equally distributes directions in the orders 0 to 7 were calculated and applied to the FDN output using the MatrixMultiplier plug-in[1]. The decoding of the signals employed the BinauralDecoder plug-in[1] that uses the MagLS approach [8, 9] on a full-sphere head-related impulse responses data set of the Neumann KU 100 dummy head [17].

For both reverberation times, a male speech recording and a transient drum sound were chosen as audio signals. The listening experiment was performed in a MUSHRA-like [18] multi stimulus comparison as in [16, 19]. In each trial, the similarity of the playback at Ambisonic orders 0 to 6 was compared to the 7^{th}-order reference. Overall, there were 2 (signals) × 2 (reverberation times) = 4 trials with 7 stimuli each.

Due to pandemic restrictions, the experiments were not conducted at the laboratory but at the individual homes of each participant using their own computer and high-quality headphones. In order to make the participation as simple as possible, we designed the experimental setup accordingly. The setup consisted of 2 components: The digital audio workstation Reaper[1] did all the audio processing and the stand-alone MUSHRA application, developed at the Institute of Electronic Music and Acoustics, controlled the entire experiment and did randomization between and within trials. The collected data was stored in a simple text file, which was subsequently sent to the examiner via email. Both components communicated via OpenSoundControl[2].

There were 10 participants in the experiment (1 female, 9 male). All of them were either master students in audio engineering or musicians.

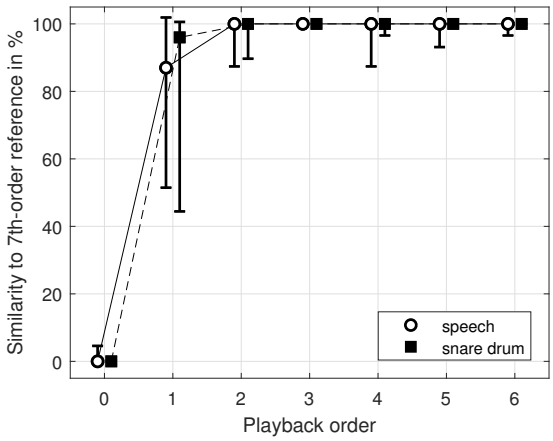
[1] freely available as part of the IEM Plug-in Suite: `https://plugins.iem.at/`

[1] `https://www.reaper.fm/`

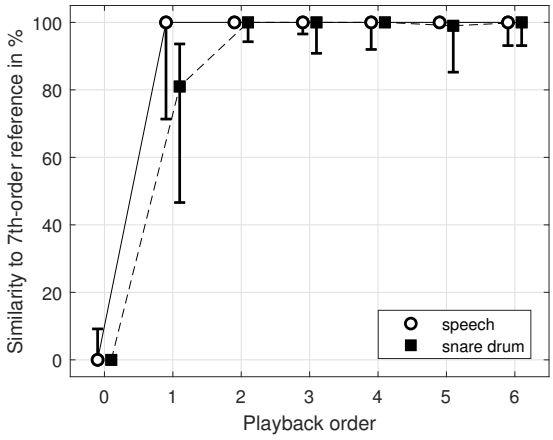
[2] `http://opensoundcontrol.org/`

Results

Figure 1 shows the median values and the corresponding confidence intervals of the perceived similarity to the 7^{th}-order reference for playback with orders 0 to 6. Independent of the reverberation length and the audio signal, 0^{th}-playback is clearly different from the reference. The listeners did not only notice the diotic playback but also reported coloration. As expected, the differences decrease with the playback order.



(a) short



(b) long

Figure 1: Medians and 95% confidence intervals of perceived similarity to 7^{th}-order reference using diffuse FDN reverb.

A Bonferroni-Holm-corrected Wilcoxon signed-rank test was used to determine the minimum required order necessary to be perceptually indistinguishable from the reference at the 5% level. Table 1 shows that for the speech signal, 1^{st}-order playback is sufficient independent of the reverberation length. The same is true for the drum signal, however only for the short reverberation. In case of the long reverberation, 2^{nd} order is necessary.

Table 1: Minimum required order to be indistinguishable from reference at 5% level with Bonferroni-Holm correction.

reverberation length	speech	drum
short	1	1
long	1	2

The results for the short reverberation agree with the design approach of the used MagLS binaural decoder: For completely diffuse sound fields, its covariance constraint should create identical order-independent output starting from 1^{st} order. For the long reverberation and the drum signal, this seems not to be the case. The reason for this is assumed to be the incomplete diffuseness of the long reverberation. In fact, the output of the feedback delay network becomes diffuse when the signal is feed through multiple times. The early, non-diffuse part is typically suppressed by subtracting the output from a second network with identical parameter but shorter reverberation time, creating a fade-in [15]. However, our setting for the fade-in time in the case of the long reverberation was chosen too short in relation to the reverberation time of the main network, resulting in too much non-diffuse sound at the beginning. The transient nature of the drum signal reveals this non-diffuse part and leads to an audible difference that cannot be compensated by the covariance filters.

Conclusion and Outlook

This contribution evaluated the perceptually required spatial resolution for binaural playback of diffuse reverberation. The reverberation was created by a feedback delay network with a reverberation time of 1.2 s and 6.7 s, respectively. The evaluation compared 0^{th}- to 6^{th}-order playback to a 7^{th}-order reference decoded to headphones using the MagLS approach [9]. For speech, the minimum required order was found to be 1 for both reverberation times, while for a transient drum sound, the required order increased to 2 in case of the long reverberation time.

The generally very low required orders perceptually verify the covariance filters in the MagLS decoder, which preserve the interchannel correlation and timbre under completely diffuse conditions even for a playback order of 1. In comparison to the results in [11] for entire impulse responses, our rendering of the diffuse part requires less spatial resolution. The increase of the required order in case of the transient sound and the long reverberation time is attributed to the short fade-in time of the feedback delay network that results in too much non-diffuse sound at the beginning of the reverberation [20, 15].

Future research could investigate a completely diffuse case by using a large number of completely decorrelated noise sources to verify our results. Moreover, it would be interesting to determine the spatial resolution of early reflections in order to find the most efficient way to render the different parts of a spatial impulse response. This is especially promising, as present findings suggest an encouragingly small number of perceptually relevant reflections [21, 22].

References

[1] A. Lindau and S. Weinzierl, "Assessing the plausibility of virtual acoustic environments," *Acta Acustica united with Acustica*, vol. 98, no. 5, pp. 804–810, 2012.

[2] S. Werner, F. Klein, A. Neidhardt, U. Sloma, C. Schneiderwind, and K. Brandenburg, "Creation of auditory augmented reality using a position-dynamic binaural synthesis system - technical components, psychoacoustic needs, and perceptual evaluation," *Applied Sciences*, vol. 11, no. 3, p. 1150, 2021.

[3] A. Lindau, H.-J. Maempel, and S. Weinzierl, "Minimum brir grid resolution for dynamic binaural synthesis," *Journal of the Acoustical Society of America*, vol. 123, no. 5, p. 3498, 2008.

[4] M. Zaunschirm, M. Frank, and F. Zotter, "Binaural rendering with measured room responses: First-order ambisonic microphone vs. dummy head," *Applied Sciences*, vol. 10, no. 5, 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/5/1631

[5] F. Zotter and M. Frank, *Ambisonics - A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, ser. Springer Topics in Signal Processing. Springer, 2019.

[6] K. Enge, M. Frank, and R. Höldrich, "Listening experiment on the plausibility of acoustic modeling in virtual reality," in *Fortschritte der Akustik-DAGA*, 2020.

[7] B. Bernschütz, A. V. Giner, C. Pörschmann, and J. Arend, "Binaural reproduction of plane waves with reduced modal order," *Acta Acustica united with Acustica*, vol. 100, no. 5, pp. 972–983, 2014.

[8] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural rendering of Ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3616–3627, 2018.

[9] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural Rendering of Ambisonic Signals via Magnitude Least Squares," in *Fortschritte der Akustik - DAGA*, Munich, March 2018.

[10] M. Zaunschirm, M. Frank, and F. Zotter, "BRIR synthesis using first-order microphone arrays," in *prepr. 9944, 144th AES Conv.*, Milano, 2018.

[11] I. Engel, C. Henry, S. V. Amengual Garí, P. W. Robinson, and L. Picinali, "Perceptual implications of different ambisonics-based methods for binaural reverberation," *The Journal of the Acoustical Society of America*, vol. 149, no. 2, pp. 895–910, 2021. [Online]. Available: https://doi.org/10.1121/10.0003437

[12] A. Lindau, L. Kosanke, and S. Weinzierl, "Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses," in *Audio Engineering Society Convention 128*. Audio Engineering Society, 2010.

[13] J. Stautner and M. Puckette, "Designing Multi-Channel Reverberators," *Computer Music Journal*, vol. 6, no. 1, pp. 52–65, 1982.

[14] J.-M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," in *90th AES Conv., prepr. 3030*, Paris, February 1991.

[15] N. Meyer-Kahlen, S. J. Schlecht, T. Lokki *et al.*, "Fade-in control for feedback delay networks," in *International Conference on Digital Audio Effects*, 2020.

[16] M. Frank and M. Brandner, "Perceptual Evaluation of Spatial Resolution in Directivity Patterns," in *Fortschritte der Akustik, DAGA*, Rostock, Mar. 2019.

[17] B. Bernschütz, "A spherical far field hrir/hrtf compilation of the neumann ku 100," in *Proceedings of the 40th Italian (AIA) annual conference on acoustics and the 39th German annual conference on acoustics (DAGA) conference on acoustics*. AIA/DAGA, 2013, pp. 592–595.

[18] ITU-R, "Bs.1534: Method for the subjective assessment of intermediate quality levels of coding systems," Tech. Rep., 2015. [Online]. Available: itu.int/rec/R-REC-BS.1534

[19] M. Frank and M. Brandner, "Perceptual Evaluation of Spatial Resolution in Directivity Patterns 2: coincident source/listener positions," in *International Conference on Spatial Audio, ICSA*, 2019.

[20] M. Blochberger, F. Zotter, and M. Frank, "Sweet area size for the envelopment of a recursive and a non-recursive diffuseness rendering approach," in *International Conference on Spatial Audio, ICSA*, 2019.

[21] S. Clapp and B. U. Seeber, "Sound Localization in Partially Updated Room Auralizations," in *Fortschritte der Akustik, DAGA*, 2016, pp. 558–560.

[22] F. Brinkmann, H. Gamper, N. Raghuvanshi, and I. Tashev, "Towards encoding perceptually salient early reflections for parametric spatial audio rendering," in *Audio Engineering Society Convention 148*. Audio Engineering Society, 2020.