Audio Engineering Society

Conference Paper

Presented at the AES International Conference on
Audio for Virtual and Augmented Reality
2020 August 17 – 19, Online

Binaural Reproduction using Bilateral Ambisonics

Zamir Ben-Hur[1], David Lou Alon[2], Ravish Mehra[2], and Boaz Rafaely[1]

[2]*School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel*
[2]*Facebook Reality Labs, Facebook, 1 Hacker Way, Menlo Park, CA 94025, USA*

Correspondence should be addressed to Zamir Ben-Hur (`zami@post.bgu.ac.il`)

ABSTRACT

Binaural reproduction plays an important role in virtual and augmented reality applications. The rendering of binaural signals using Spherical-Harmonics (SH) representation gives the flexibility to control the reproduced binaural signals, by using algorithms that operate directly in the SH domain. However, in most practical cases, the binaural reproduction is order-limited, which introduces truncation error that has a detrimental effect on the perception of the reproduced signals. A recent study showed that pre-processing of the Head-Related Transfer Function (HRTF) by ear-alignment reduces its effective SH order. In this paper, a method to incorporate the ear-aligned HRTF into the binaural reproduction process using a new Ambisonics representation of the sound field formulated at the two ears, denoted here as Bilateral Ambisonics, is presented. Application of this method yields a significant improvement in the perceived audio quality of order-limited binaural signals.

1 Introduction

Binaural reproduction technology provides the listener with the sensation of being present in the 3D audio scene. Binaural signals can be synthesized in the Spherical-Harmonics (SH) domain, using SH representations of the sound field and the Head-Related Transfer Function (HRTF) [1, 2]. Such processes give the flexibility to control the reproduced binaural signals, by manipulating the sound field or the HRTF using algorithms that operate directly in the SH domain [3]. However, in most practical cases the binaural reproduction is order-limited, which introduces truncation error. This error is caused by the order truncation of the HRTF [4], leading to significant artifacts, both in space and in frequency, which have a detrimental effect on perceived audio quality in general, and, in particular, on the perceived timbre, localization, externalization, source width and stability of the virtual sound sources [4, 5, 6].

Several methods which operate as post-processing on the binaural signals have been proposed to reduce these errors [4, 7, 8, 9]. In particular, pre-processing of the HRTF has also been shown to reduce its effective SH order [10], which may potentially reduce the truncation error. Nevertheless, even with current methods, a low-order binaural signal is typically perceived as significantly different from a high-order reference [11]. High quality binaural reproduction using a low SH order therefore remains an open problem.

In this paper, a method to incorporate the ear-alignment technique that was recently developed for pre-processing of HRTFs [12] with a reduced SH order into the binaural reproduction process is presented. The proposed binaural reproduction method is based on convolving an ear-aligned HRTF with a new Bilateral Ambisonics representation of the sound field formulated around the listener's ears, as detailed in the next section.

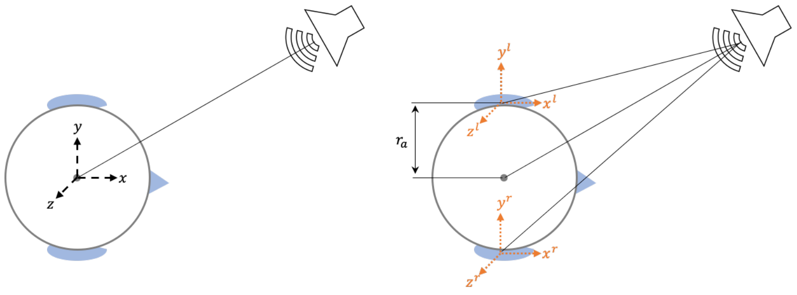
2 Bilateral Ambisonics Reproduction

A binaural signal can be calculated in the SH domain using the standard Ambisonics reproduction by [3]:

$$p^l(k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} [\tilde{a}_{nm}(k)]^* h^l_{nm}(k), \qquad (1)$$

where k is the wave number, $p^l(k)$ is the pressure at the left ear, $a_{nm}(k)$ is the spherical Fourier transform (SFT) of the Plane-Wave (PW) density function, $a(k,\Omega)$, which encodes the directional properties of the sound field [13], and $\tilde{a}_{nm}(k)$ is the SFT of $a^*(k,\Omega)$. $[\cdot]^*$ denotes the complex conjugate, Ω is the spatial angle, and $h^l_{nm}(k)$ are the SH coefficients of the left ear HRTF, which can be computed by applying the SFT to the HRTF. The superscript l denotes the left ear (the formulation for the right ear can be similarly expressed). N denotes the SH order, which is limited by the available orders of the sound-field and the HRTF [14]. In practice, $\tilde{a}_{nm}(k)$, which is the Ambisonics signal, can be derived from spherical microphone array recordings, and its order will be limited by the number of microphones [15]. However, the HRTF is inherently of high spatial order [16]. Therefore, due to (1), the HRTF will be truncated to the order of the sound-field. The impact of this order truncation is in both the spectral and the spatial domains, and the corrupted binaural signal leads to degradation in the perceived spatial sound quality [4, 5].

Recently, several methods have been proposed to pre-process the HRTF as a prior stage of integrating the HRTF in the SH domain. A few prominent examples employ time-alignment [2, 17], minimum-phase representation [18], directional equalization [19], or ear-alignment [12]. A common theme of these approaches is their focus on the linear-phase HRTF component, which is largely responsible for the higher orders found in the true HRTF. A drawback of these methods lies in the fact that the pre-processed HRTF cannot participate in real time in the binaural signal computation, as a result of its interaction with the sound field. In this paper, we propose a framework for the computation of the binaural signals directly at the position of the ear, using the ear-aligned HRTFs to reduce their effective SH order [12]. The framework exploits the reduced-order HRTF, leading to more accurate binaural signals with low-order reproduction. The framework uses a novel representation of the sound field denoted in this paper as Bilateral Ambisonics. This representation is composed of two Ambisonics signals, each defined around one of the two ears, rather than around the center of the head, as in the standard Ambisonics. The Bilateral



(a) Standard coordinate system.

(b) Bilateral coordinate system.

Fig. 1: Schematic of the two coordinate systems: (a) the standard coordinate system, where the origin is at the center of the head, and (b) the Bilateral coordinate system, where the origin is located at the ear.

Ambisonics reproduction is formulated as:

$$p^l(k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} [\tilde{a}^l_{nm}(k)]^* h^{al}_{nm}(k), \qquad (2)$$

where $\tilde{a}^l_{nm}(k)$ and $h^{al}_{nm}(k)$ are the SH coefficients of the left ear Ambisonics signal, defined around the position of the left ear (in contrast to $\tilde{a}_{nm}(k)$, which is the standard Ambisonics signal, defined around the position of the center of the head) and the left ear-aligned HRTF, respectively. A similar computation should also be performed for the right ear. In the schematic in Fig. 1 the two coordinate systems are shown. Figure 1(a) describes the coordinate system used with the standard Ambisonics signals, $\tilde{a}_{nm}(k)$, for both ears, where the binaural signals are calculated from Eq. (1). Figure 1(b), on the other hand, describes the coordinate system used with the Bilateral Ambisonics signals, $\tilde{a}^l_{nm}(k)$ and $\tilde{a}^r_{nm}(k)$ for the left and right ears, respectively. Here, the binaural signals are calculated from Eq. (2).

Ear-alignment of the HRTF is performed by translating the origin of the free-field component of the HRTF from the center of the head to the position of the ear. Thus, the SH coefficients of the ear-aligned HRTF can be readily computed from a typical HRTF [12]. However, the SH coefficients of the sound field at the position of the ears are often not available. This is not the case when the sound field is generated using numerical simulations, since the Bilateral Ambisonics signals can be directly simulated at the two ear positions. Likewise, in the case of binaural reproduction based on sound field measurements, the Bilateral Ambisonics signals can be derived from two microphone array measurements at the positions of the ears, or estimated from standard recordings at the center of the head, e.g. by sound field translation [20, 21, 22, 23]. Theoretically, if a low-order Ambisonics signal is given directly at the position of the ear, the Bilateral Ambisonics based binaural signal, as in Eq. (2), may potentially be more accurate than the standard reproduction (as in Eq. (1)), because of the lower-order nature of the ear-aligned HRTF, compared to a standard HRTF.

3 Subjective Evaluation of the Proposed Method

To evaluate the performance of the proposed Bilateral Ambisonics reproduction approach, a listening experiment was conducted. The experiment compared the proposed method with the standard Ambisonics reproduction with equalization and tapering as suggested in [7]. This configuration was selected as it was reported to reduce reproduction errors for low-order binaural signals, while offering similar perceptual performance to other previously suggested methods [11].

3.1 Test Signals

The test signals were computed using the standard Ambisonics reproduction in the SH domain, as in Eq. (1), with equalization and tapering [7], and using the suggested Bilateral Ambisonics reproduction, as in Eq. (2).

The HRTF set used for the computation of the test signals is the measured HRTF from the Cologne HRTF database for the Neumann KU100 dummy head [24]. The ear-aligned HRTF was computed using nominal parameters as suggested in [12].

Room impulse responses were simulated using the image method [25]. The room was cuboid in shape, of size $15.5 \times 9.8 \times 7.5$ m with reverberation time, $T_{60} = 0.96$ s and critical distance of 1.94 m. The locations of a simulated spherical microphone array and an omnidirectional source were $(x, y, z) = (5, 2.5, 1.7)$ m and $(7.25, 3.8, 1.7)$ m, respectively. It can be seen that the latter was distance 2.6 m from the former, where the angular displacement was 30° to the left, relative to the HRTF coordinate system. In the case of the Bilateral Ambisonics reproduction the two arrays were located by the ears, see Fig. 1, where $r_a = 8.75$ cm. The Ambisonics signals, $a_{nm}(k)$ and $a^{l/r}_{nm}(k)$, were simulated at the desired SH order, directly in the SH domain, such that no spatial aliasing is introduced.

Three SH orders were used for both the standard and the Bilateral Ambisonics reproductions, $N = 1$, 2 and 6. These were selected based on a preliminary informal listening test. The reference signal was of order $N = 41$, rendered with the standard Ambisonics reproduction. Two different audio source signals were played to the subjects: male speech in English and castanets. All signals were convolved with matching headphone compensation filters, taken from the Cologne database [24], which were measured on the Neumann KU100 dummy head, and their loudness levels were equalized [26].

3.2 Method

The participants in the listening tests had previous experience with similar experiments, and had no known hearing impairments. 13 males and 2 females aged 27-49 participated in both tests (castanets and male speech), which were conducted in accordance with the multiple stimuli with hidden reference and anchor (MUSHRA) protocol [27]. In each test the two reproduction methods were examined with the three SH orders N, as described above. Thus, with the addition of the hidden reference, there were 7 test signals, see Fig. 2. In MUSHRA, the subjects rank the similarity between the test and reference signals between 100 (indistinguishable) and 0 on the basis of several parameters: spectral artifacts, spatial artifacts, added noise or time varying artifacts. Since head tracking was unavailable, the subjects were told to remain stationary as much as possible. At the beginning of the experiment, a training session was conducted with the aims of learning how to use the test equipment and the grading scale, and becoming familiar with all the test signals and their quality level ranges.

3.3 Results

The results of all the tests for 13 out of the 15 subjects are shown in Fig. 2. The results of 2 participants were disqualified as they were not able to identify the hidden reference adequately (ranked below 80 in at least one test). The proposed Bilateral Ambisonics method clearly demonstrated perceptual improvement compared with the standard Ambisonics method. The statistical significance of the results were determined with a three-factorial ANOVA with the factors "Reproduction" (Standard, Bilateral), "Audio source signal" (Castanets, Speech) and "SH order" ($N = 1, 2, 6$), paired with a Tukey-Kramer post hoc test at a confidence level of 95%, which was used to determine the statistical significance of the results [27, 28]. Statistical significance was found for all factors: $F_{(2,146)} = 60.7$, $F_{(1,146)} = 308.59$ and $F_{(1,146)} = 43.26$ for "SH order", "Reproduction" and "Audio source signal", respectively, with $p_{\text{val}} < 0.001$ for all of them. Pairwise comparisons between the test signals for the "Reproduction" factor reveal that with the two audio source signals 1st order Bilateral Ambisonics reproduction achieved statistically significantly higher scores compared to 1st and 2nd order standard Ambisonics reproduction ($p_{\text{val}} < 0.001$). In addition, the median scores for the 1st order Bilateral Ambisonics reproduction are higher compared to the 6th order standard Ambisonics reproduction (70 compared to 65 for the castanets, and 100 compared to 90 for the speech). However, these differences are not statistically significant ($p_{\text{val}} > 0.3$).

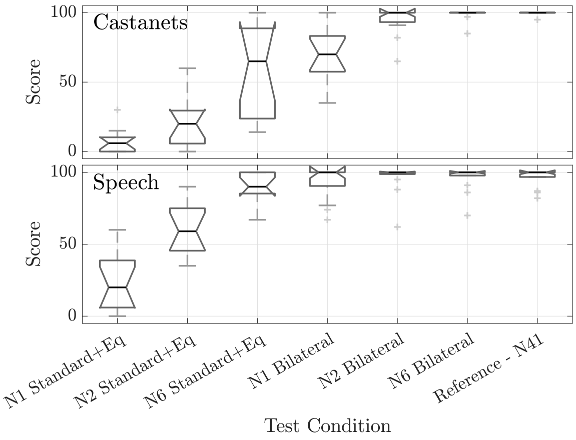


Fig. 2: Results of the listening tests: upper plot - castanets and lower plot - male speech. The different boxes in the box plot represent the test conditions listed on the x axis. The notches in the boxes indicate the 95% confidence interval, while the box bounds the ratio between the interquartile range and the median. Tukey-style whiskers have been added to indicate a maximum of of 1.5×IQR beyond the box [29].

The results show that for the Bilateral Ambisonics case the test signals were indistinguishable from the order $N = 41$ reference ($p_{\text{val}} > 0.98$), with the exception of the castanets test with SH order $N = 1$ ($p_{\text{val}} = 0.001$). In contrast, for the standard Ambisonics case the test signals were clearly distinguishable from the reference for SH orders $N = 1$ and 2, while the castanets signal was noticeably different from the reference even when its SH order was $N = 6$.

4 Conclusion

The current paper presented a new method for binaural reproduction in the SH domain based on Bilateral Ambisonics and ear-aligned HRTFs. Listening test results indicate that the proposed Bilateral Ambisonics reproduction with an order as low as $N = 1$ has significant perceptual benefits over the standard Ambisonics reproduction. Moreover, 2nd order Bilateral Ambisonics reproduction achieved similar scores to those of the high-order reference.

Acknowledgment

This research was supported by Facebook Reality Labs.

References

[1] Blauert, J., *Spatial hearing: the psychophysics of human sound localization*, MIT press, 1997.

[2] Evans, M. J., Angus, J. A., and Tew, A. I., "Analyzing head-related transfer function measurements using surface spherical harmonics," *The Journal of the Acoustical Society of America*, 104(4), pp. 2400–2411, 1998.

[3] Rafaely, B. and Avni, A., "Interaural cross correlation in a sound field represented by spherical harmonics," *The Journal of the Acoustical Society of America*, 127(2), pp. 823–828, 2010.

[4] Ben-Hur, Z., Brinkmann, F., Sheaffer, J., Weinzierl, S., and Rafaely, B., "Spectral equalization in binaural signals represented by order-truncated spherical harmonics," *The Journal of the Acoustical Society of America*, 141(6), pp. 4087–4096, 2017.

[5] Avni, A., Ahrens, J., Geier, M., Spors, S., Wierstorf, H., and Rafaely, B., "Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution," *The Journal of the Acoustical Society of America*, 133(5), pp. 2711–2721, 2013.

[6] Ben-Hur, Z., Alon, D. L., Rafaely, B., and Mehra, R., "Loudness stability of binaural sound with spherical harmonic representation of sparse head-related transfer functions," *EURASIP Journal on Audio, Speech, and Music Processing*, 2019(1), p. 5, 2019, ISSN 1687-4722, doi:10.1186/s13636-019-0148-x.

[7] Hold, C., Gamper, H., Pulkki, V., Raghuvanshi, N., and Tashev, I. J., "Improving binaural ambisonics decoding by spherical harmonics domain tapering and coloration compensation," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 261–265, IEEE, 2019.

[8] Schörkhuber, C., Zaunschirm, M., and Höldrich, R., "Binaural rendering of ambisonic signals via magnitude least squares," in *Proceedings of the DAGA*, volume 44, pp. 339–342, 2018.

[9] Alon, D. L., Sheaffer, J., and Rafaely, B., "Plane-wave decomposition with aliasing cancellation for binaural sound reproduction," in *Audio Engineering Society Convention 139*, Audio Engineering Society, 2015.

[10] Brinkmann, F. and Weinzierl, S., "Comparison of Head-Related Transfer Functions Pre-Processing Techniques for Spherical Harmonics Decomposition," in *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*, Audio Engineering Society, 2018.

[11] Lübeck, T., *Perceptual Evaluation of Mitigation Approaches of Errors due to Spatial Undersampling in Binaural Renderings of Spherical Microphone Array Data*, Master's thesis, Chalmers University of Technology, Gothenburg, Sweden, 2019.

[12] Ben-Hur, Z., Alon, D. L., Mehra, R., and Rafaely, B., "Efficient Representation and Sparse Sampling of Head-Related Transfer Functions Using Phase-Correction Based on Ear Alignment," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(12), pp. 2249–2262, 2019.

[13] Rafaely, B., "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *The Journal of the Acoustical Society of America*, 116(4), pp. 2149–2157, 2004.

[14] Ben-Hur, Z., Sheaffer, J., and Rafaely, B., "Joint sampling theory and subjective investigation of plane-wave and spherical harmonics formulations for binaural reproduction," *Applied Acoustics*, 134, pp. 138–144, 2018.

[15] Rafaely, B., *Fundamentals of Spherical Array Processing*, volume 8, Springer, 2015.

[16] Zhang, W., Abhayapala, T. D., Kennedy, R. A., and Duraiswami, R., "Insights into head-related transfer function: Spatial dimensionality and continuous representation," *The Journal of the Acoustical Society of America*, 127(4), pp. 2347–2357, 2010.

[17] Zaunschirm, M., Schörkhuber, C., and Höldrich, R., "Binaural rendering of Ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *The Journal of the Acoustical Society of America*, 143(6), pp. 3616–3627, 2018.

[18] Romigh, G. D., Brungart, D. S., Stern, R. M., and Simpson, B. D., "Efficient real spherical harmonic representation of head-related transfer functions," *IEEE Journal of Selected Topics in Signal Processing*, 9(5), pp. 921–930, 2015.

[19] Pörschmann, C., Arend, J. M., and Brinkmann, F., "Directional Equalization of Sparse Head-Related Transfer Function Sets for Spatial Upsampling," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(6), pp. 1060–1071, 2019.

[20] Duraiswami, R., Li, Z., Zotkin, D. N., Grassi, E., and Gumerov, N. A., "Plane-wave decomposition analysis for spherical microphone arrays," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005.*, pp. 150–153, IEEE, 2005.

[21] Pihlajamaki, T. and Pulkki, V., "Synthesis of complex sound scenes with transformation of recorded spatial sound in virtual reality," *Journal of the Audio Engineering Society*, 63(7/8), pp. 542–551, 2015.

[22] Birnie, L., Abhayapala, T., Samarasinghe, P., and Tourbabin, V., "Sound Field Translation Methods for Binaural Reproduction," in *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 140–144, IEEE, 2019.

[23] Kentgens, M., Behler, A., and Jax, P., "Translation of a Higher Order Ambisonics Sound Scene Based on Parametric Decomposition," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 151–155, IEEE, 2020.

[24] Bernschütz, B., "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*, p. 29, 2013.

[25] Allen, J. B. and Berkley, D. A., "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, 65(4), pp. 943–950, 1979.

[26] ITU-R-BS, "1770-4: Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level," 2015.

[27] ITU-R-BS, "1534-3: Method for the subjective assessment of intermediate quality level of audio systems," 2015.

[28] Breebaart, J., "Evaluation of statistical inference tests applied to subjective audio quality data with small sample size," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 23(5), pp. 887–897, 2015.

[29] Tukey, J. W., *Exploratory data analysis*, volume 2, Addison-Wesley, Reading, Massachusetts, pp. 39–42, 1977.