Towards a perceptually optimal bias factor for directional bias equalisation of binaural ambisonic rendering

Digital Object Identifier (DOI):
10.25836/sasp.2019.08

Published In:
Proceedings of the EAA Spatial Audio Signal Processing symposium

Towards a perceptually optimal bias factor for directional bias equalisation of binaural ambisonic rendering

Thomas Mckenzie, Damian Murphy, Gavin Kearney

TOWARDS A PERCEPTUALLY OPTIMAL BIAS FACTOR FOR DIRECTIONAL BIAS EQUALISATION OF BINAURAL AMBISONIC RENDERING

Thomas McKenzie, Damian T. Murphy, Gavin Kearney
AudioLab, Communication Technologies Research Group,
Department of Electronic Engineering, University of York,
York, YO10 5DD, UK
thomas.mckenzie@york.ac.uk

ABSTRACT

In virtual reality applications, where Ambisonic audio is presented to the user binaurally (over headphones) in conjunction with a head-mounted display, retained immersion relies on congruence between the auditory and visual experiences. Therefore frontal accuracy of binaural Ambisonic audio is an area of interest. A previously introduced method for improving the frequency reproduction of Ambisonic binaural rendering of a specific direction, called directional bias equalisation, is here applied to higher order Ambisonics and evaluated both numerically and perceptually. This paper attempts to obtain the optimal amount of directional bias, to find the best compromise between improved frontal reproduction and reduced lateral reproduction accuracy.

1. INTRODUCTION

Ambisonics is a method of generating, capturing and rendering two or three-dimensional sound fields, first developed by Michael Gerzon in the 1970s [1]. It has enjoyed a recent resurgence in popularity due to virtual reality applications, where Ambisonic audio can be presented to the user binaurally [2, 3] in conjunction with a head-mounted display, due to the relative ease of Ambisonic soundfield rotation. In virtual reality scenarios however, it is imperative to maximise the coherence between audio in the frontal direction and visuals in order to maintain immersion.

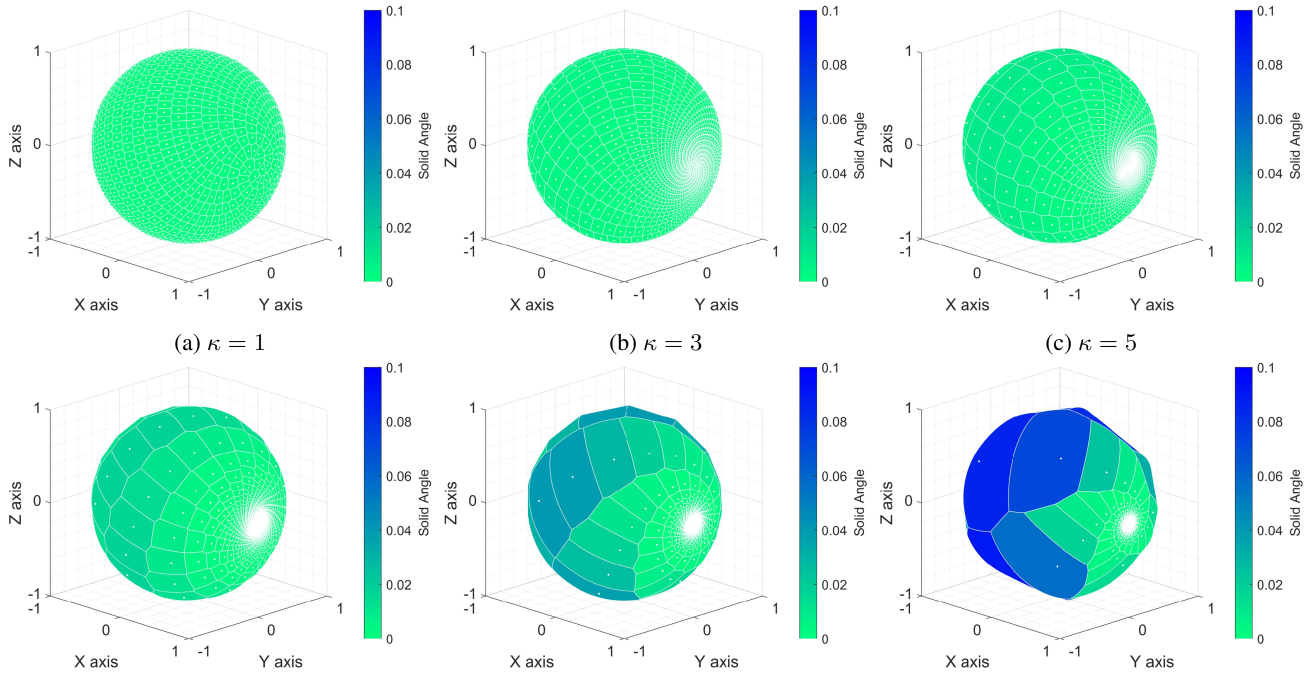
Ambisonic reproduction can theoretically be perfect in the centre of the loudspeaker array for frequencies up to the 'spatial aliasing frequency', f_{alias}. At frequencies above f_{alias} however, the limited spatial accuracy of reproducing a physical sound field with a finite number of transducers produces artefacts such as localisation blur, reduced lateralisation and comb filtering. Increasing the order of Ambisonics delivers a more accurate soundfield reconstruction, but requires a greater number of channels and convolutions in binaural rendering. One approach for improving spectral reproduction of binaural Ambisonic rendering is diffuse-field equalisation [4]. In a previous study, the authors applied this technique to virtual loudspeaker binaural Ambisonic decoders, which was shown to improve both the spectral response over the sphere and predicted median plane elevation localisation. However, it was evident there still exists a definite and perceivable difference in timbre between diffuse-field equalised binaural Ambisonic rendering and HRTF convolution, even at 5th order Ambisonics.

Altering the diffuse-field equalisation method by concentrating the equalisation in one specific direction has been shown to be possible by creating a hybrid of free-field and diffuse-field equalisation. Reproduction in the desired direction can then be made more accurate at the expense of other directions, such that an infinite directional bias can theoretically produce binaural Ambisonic audio equivalent to HRTF convolution. This is referred to as directional bias equalisation (DBE) [5]. DBE is a pre-processing stage that can be implemented offline.

In the initial paper detailing the DBE method, evaluation focused on application to 1st order Ambisonics with only a numerical analysis presented. This paper follows on from the first paper, extending the evaluation of the technique to higher-order Ambisonics, with both a numerical and perceptual evaluation. Here 1st, 3rd and 5th order Ambisonics are investigated, with loudspeaker configurations comprising 6, 26 and 50 loudspeakers respectively, arranged in Lebedev grids [6] (for exact vertices, see [7]). The 6 values of the bias factor κ used in this paper are: $\kappa = 1, 3, 5, 9, 17$ and 33. Though DBE can increase the fidelity of any desired direction, this paper will focus on frontal bias at a direction of $(\theta, \phi) = (0°, 0°)$, where θ and ϕ denote azimuth and elevation, respectively.

In this study Ambisonics was rendered binaurally with mode-matching pseudo-inverse decoding using the Politis Ambisonic library [8]. Three-dimensional full normalisation (N3D) and Ambisonic channel number (ACN) ordering was used throughout. Dual-band decoding was implemented with Max r_E weighting [9, 10] at frequen-

doi:10.25836/sasp.2019.08



(a) $\kappa = 1$ (b) $\kappa = 3$ (c) $\kappa = 5$

(d) $\kappa = 9$ (e) $\kappa = 17$ (f) $\kappa = 33$

Figure 1: Voronoi spheres demonstrating the directionally biased quadratures used in the first stage of DBE for the 6 values of κ used in this study.

cies above the spatial aliasing frequency, which in this study was approximated according to [11, 12] as 670 Hz, 1870 Hz and 3070 Hz for 1st, 3rd and 5th order Ambisonics, respectively, calculated with a speed of sound of 343 m/s and radius of the listening area as 9 cm (the approximate radius of the Neumann KU 100 dummy head). All head-related impulse responses (HRIRs) used in this study were from the Bernschütz Neumann KU 100 database [13]. All computation was carried out offline in MATLAB version 9.3.0 - R2017b. All audio used was of 24-bit depth and 48 kHz sample rate.

2. DIRECTIONAL BIAS EQUALISATION

DBE is a two stage equalisation process. For a detailed explanation of the DBE method, the reader is directed to the original paper [5]. Here the method will be briefly summarised. Firstly, a directionally biased quadrature (DBQ) RMS response is calculated from the root-mean-square (RMS) of the magnitude responses of a large number of binaural Ambisonic renders, taken at locations over the sphere. The quadrature is directionally biased by skewing the z axis values according to κ such that:

$$z_\beta = \kappa(z_\alpha + 1) - 1 \tag{1}$$

The cartesian coordinates are then converted back to spherical coordinates and rotated to the direction of bias. In this study, Fibonacci quadrature with 1000 points is used in DBQ RMS calculations, due to its relative even distribution of points. The Voronoi sphere plots of the 6 values of κ used in this paper are shown in Figure 1. With no directional bias (see Figure 1a), points are evenly distributed over the sphere, which when equalised produces an even RMS response of the binaural Ambisonic loudspeaker configuration, theoretically equivalent to diffuse-field equalisation [4].

The second stage of DBE is a directional bias HRTF equalisation. In this study the direction of bias is chosen as directly in front: $(\theta, \phi) = (0°, 0°)$. As κ increases, the gain of the frontal bias HRTF g_β increases such that:

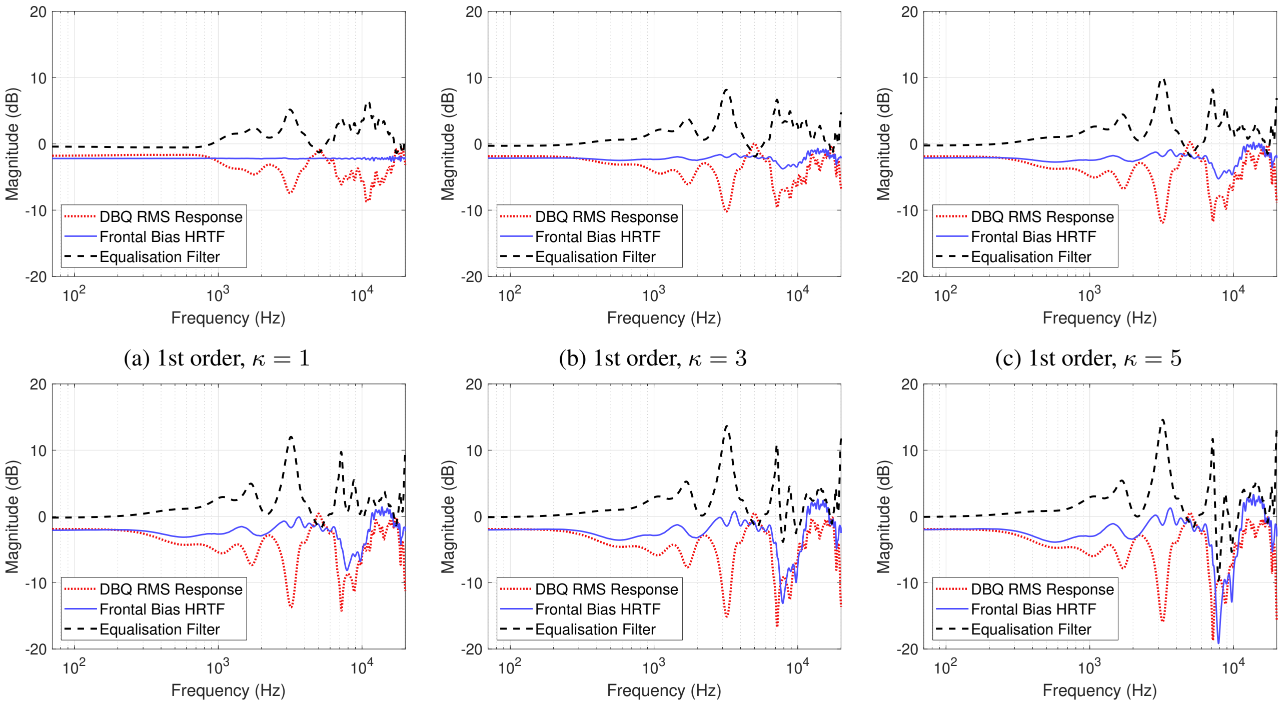
$$g_\beta = 1 - e^{-\frac{\kappa - 1}{10}} \tag{2}$$

The frontal bias HRTF filter is then convolved with the DBQ RMS equalisation resulting in the final DBE filters. Figure 2 presents the DBQ RMS responses, frontal bias HRTFs, and resulting overall equalisation filters for 1st order, with varying κ. To conserve space, 3rd and 5th order plots have been omitted. As κ increases, the DBQ response closer resembles a frontal Ambisonic render, and the frontal bias HRTF closer resembles a frontal HRTF.

3. NUMERICAL EVALUATION

To assess the effect of varying directional bias on the spectral accuracy of binaural Ambisonic signals, a perceptually motivated fast Fourier transfer (FFT) based spectral difference (PSD) model was used that takes into account various features of human auditory perception [14]. The PSD model weights input signals using ISO 226 equal loudness contours [15] to account for the frequency-varying sensitivity of human hearing, uses a sone scale to account for the loudness-varying sensitivity of human hearing, and equivalent rectangular bandwidth weightings to address how the

doi:10.25836/sasp.2019.08



(a) 1st order, $\kappa = 1$ (b) 1st order, $\kappa = 3$ (c) 1st order, $\kappa = 5$

(d) 1st order, $\kappa = 9$ (e) 1st order, $\kappa = 17$ (f) 1st order, $\kappa = 33$

Figure 2: Directionally biased quadrature responses, frontal bias HRTF equalisation and resulting overall equalisation filters for 1st order Ambisonics with varying κ.

linearly spaced samples of an FFT do not accurately represent the approximately logarithmic frequency sensitivity of the inner ear.

PSD between binaural Ambisonic renders and HRIRs was calculated for all 3 tested orders of Ambisonics over 16,020 locations on the sphere, distributed using a 2° Gaussian grid. Figure 3 shows the solid angle weighted PSD value for each value of κ for all tested orders of Ambisonics, with whiskers denoting the maximum and minimum PSD values. It is clear that increasing κ reduces the minimum PSD value whilst increasing the maximum PSD value. At $\kappa = 33$, the original HRTF is almost perfectly reconstructed, even with 1st order Ambisonics. However an interesting trend appears to suggest that with increased Ambisonic order, the value of κ needs to be greater to achieve a similar minimum PSD. To observe how PSD changes over the sphere with varying κ, 1st order PSD plots are presented in Figure 4 (to reduce the overall amount of figures, 3rd and 5th order plots were omitted). This shows how increasing the value of κ produces an improvement in spectral accuracy for the frontal direction with a reduction in lateral accuracy, which is in line with expectations. 3rd and 5th order renders also follow this pattern.

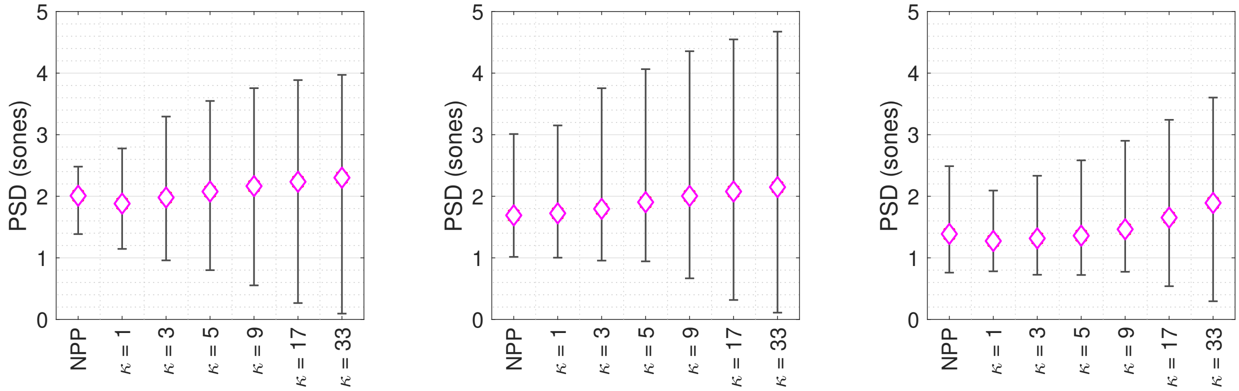
4. PERCEPTUAL EVALUATION

To assess the perceptual effect of varying κ, listening tests were conducted using both simple and complex acoustic scenes. Tests followed the multiple stimulus with hidden reference and anchors (MUSHRA) paradigm, ITU-R BS.1534-3 [16], and were conducted in a quiet listening room using a single set of Sennheiser HD 650 circum-aural headphones and an Apple Macbook Pro with a Fireface 400 audio interface, which has software controlled input and output levels. The headphones were equalised using a Neumann KU 100 from the RMS average of 11 impulse response measurements collected using Farina's swept sine technique [17]. Inverse filtering was achieved using Kirkeby and Nelson's least-mean-square regularization method [18], with one octave smoothing implemented using the complex smoothing approach of [19] and an inversion range of 5 Hz–4 kHz. In-band and out-band regularization of 25 dB and −2 dB respectively was used to avoid narrow peaks in the inverse filter, which are more noticeable than notches [20]. 20 experienced listeners participated, aged between 22 and 41, with no reported knowledge of any hearing impairments.

4.1 Test Methodology

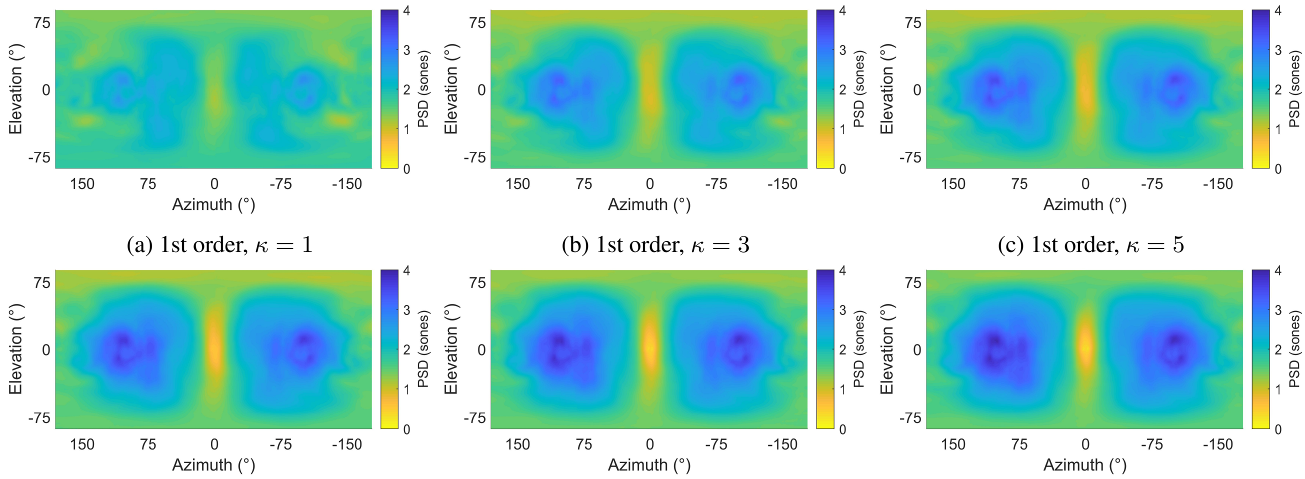
The simple scenes test comprised of a single pink noise source. Two sound source locations were used: directly in front of the listener at $(\theta, \phi) = (0°, 0°)$, and directly to the left of the listener at $(\theta, \phi) = (90°, 0°)$. The reference was a direct HRIR convolution, and low and mid anchors were the reference low-passed at 3.5 kHz and 7 kHz, respectively.

The complex scene was simulated by mixing a pink noise burst with a diffuse soundscape. The noise burst consisted of 0.5s burst followed by 0.5s of silence panned directly in front of the listener. The diffuse soundscape was synthesised from 24 excerpts of a monophonic sound scene



(a) 1st order (b) 3rd order (c) 5th order

Figure 3: Solid angle weighted perceptual spectral difference between 16,020 binaural Ambisonic renders and HRIRs with varying κ, for 1st, 3rd and 5th order Ambisonics. Whiskers denote maximum and minimum PSD values and NPP denotes no pre-processing.



(a) 1st order, $\kappa = 1$ (b) 1st order, $\kappa = 3$ (c) 1st order, $\kappa = 5$

(d) 1st order, $\kappa = 9$ (e) 1st order, $\kappa = 17$ (f) 1st order, $\kappa = 33$

Figure 4: Perceptual spectral difference between 1st-order binaural Ambisonic renders and HRIRs over the sphere with varying κ.

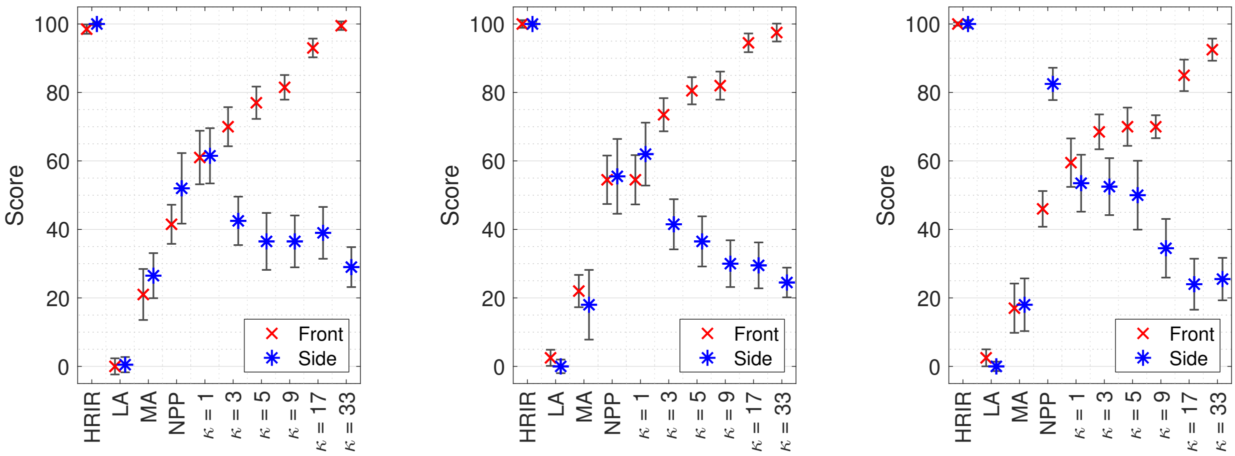
recording of a train station [21], each 5s long. The sound scene excerpts were panned to the vertices of a spherical 24 pt. T-design quadrature [22], to ensure minimal overlap between virtual loudspeaker positions in the binaural decoders and the sound sources in the complex scene. The frontal noise was set as 3 dB RMS louder than the diffuse soundscape to approximate a centre of attention. The reference comprised of a sum of direct HRIR convolutions and the anchor a 0th order Ambisonic render. All test scenarios were repeated once.

4.2 Results

Results data was tested for normality using the one-sample Kolmogorov-Smirnov test, which showed all data as non-normal. Results were therefore analysed using non-parametric statistics. Figure 5 presents the simple scene median scores for orders 1st, 3rd and 5th order Ambisonics with non-parametric 95% confidence intervals [23]. A Friedman's ANOVA, conducted on simple scene data from all orders and sound source locations, confirmed that changing the value of κ had a statistically significant effect on the perceived similarity to the HRTF reference ($\chi^2(6) = 27.22, p < 0.01$). The results support the theory that increasing κ improves the perceived similarity to the HRTF reference for the frontal stimuli for all 3 tested orders of Ambisonics, and reduces the similarity for the lateral stimuli. This shows that DBE performs as expected with simple scenes.

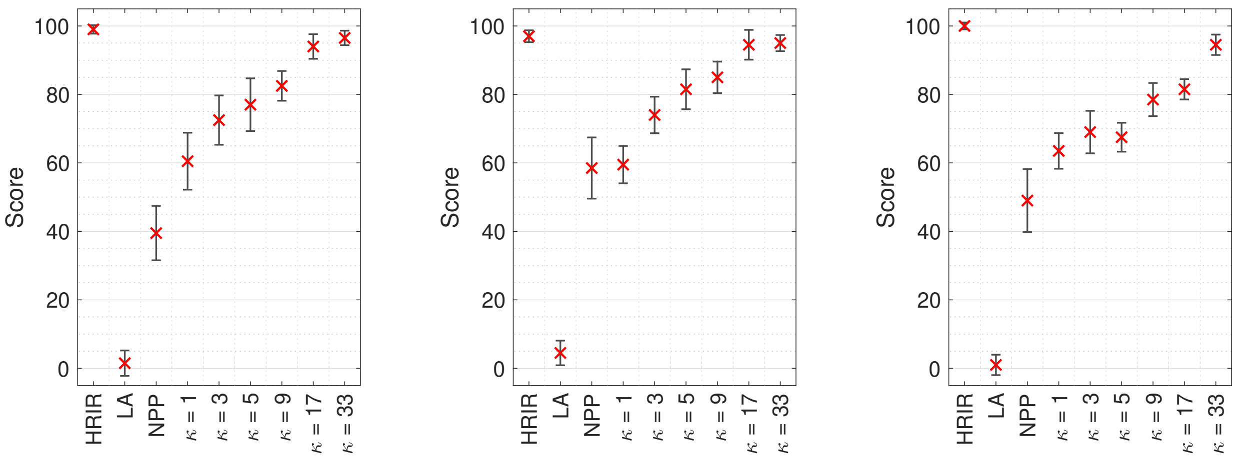
Figure 6 presents the complex scene median scores for 1st, 3rd and 5th order Ambisonics with non-parametric 95% confidence intervals [23]. A Friedman's ANOVA, conducted on complex scene data from all orders, again confirmed that changing the value of κ had a statistically significant effect on the perceived similarity to the HRTF reference ($\chi^2(6) = 383.47, p < 0.01$). Interestingly, the results suggest that the diffuse sound was essentially ignored, as results for the complex scene are similar to the frontal stimuli in the simple scene, with increasing κ producing a higher perceived similarity to the HRTF references for all 3 tested orders of Ambisonics.

doi:10.25836/sasp.2019.08



(a) 1st order (b) 3rd order (c) 5th order

Figure 5: Median scores of the simple scene tests with non-parametric 95% confidence intervals. Scores indicate perceived similarity to the HRIR reference. LA, MA and NPP denote low anchor, medium anchor and no pre-processing, respectively.



(a) 1st order (b) 3rd order (c) 5th order

Figure 6: Median scores of the complex scene tests with non-parametric 95% confidence intervals. Scores indicate perceived similarity to the HRIR reference. LA and NPP denote low anchor and no pre-processing, respectively.

5. DISCUSSION AND CONCLUSIONS

This paper has presented a numerical and perceptual evaluation of directional bias equalisation for 1st, 3rd and 5th order binaural Ambisonic rendering. It has shown how, as directional bias increases, the spectral performance in the direction of bias improves, but to the detriment of other directions. However when the scene is complex, with a frontal main source and additional diffuse sources, increased frontal bias still improves the performance. This suggests that, if the main sound source is at the front, one can afford to increase κ without greatly reducing the perceived quality in lateral directions. However, if the stimuli is more diffuse, the value of κ should be reduced.

Future work will look at combining with other pre-processing techniques for improving binaural Ambisonic reproduction, including ILD optimisation [24] and time alignment [25–27]. As both have been shown to improve spectral reproduction, this could mean that a lower value of κ is required to get perceptually equivalence to HRIR convolution in the direciton of bias. Additionally, if the dominant direction of the signal can be estimated (for example using a method such as directional audio coding [28]), the equalisation could be performed specifically for the direction of arrival, such as other signal dependent Ambisonic decoding methods [29, 30].

6. REFERENCES

[1] M. A. Gerzon, “Periphony: with-height sound reproduction,” *Journal of the Audio Engineering Society*, vol. 21, no. 1, 1973.

[2] J.-M. Jot, S. Wardle, and V. Larcher, “Approaches to binaural synthesis,” in *105th Convention of the Audio Engineering Society*, vol. 105, 1998.

[3] M. Noisternig, A. Sontacchi, T. Musil, and R. Höldrich, “A 3D Ambisonic based binaural sound reproduction system,” in *AES 24th International Conference on Multichannel Audio*, 2003.

[4] T. McKenzie, D. Murphy, and G. Kearney, "Diffuse-field equalisation of binaural Ambisonic rendering," *Applied Sciences*, vol. 8, no. 10, 2018.

[5] T. McKenzie, D. Murphy, and G. Kearney, "Directional bias equalisation of first-order binaural Ambisonic rendering," in *AES Conference on Audio for Virtual and Augmented Reality*, 2018.

[6] V. I. Lebedev, "Quadratures on a sphere," *USSR Computational Mathematics and Mathematical Physics*, vol. 16, no. 2, pp. 10–24, 1976.

[7] J. Burkardt, "SPHERE_LEBEDEV_RULE - Quadrature Rules for the Unit Sphere." 2013.

[8] A. Politis, *Microphone array processing for parametric spatial audio techniques*. PhD thesis, Aalto University, 2016.

[9] M. A. Gerzon and G. J. Barton, "Ambisonic decoders for HDTV," in *92nd Convention of the Audio Engineering Society*, 1992. Preprint 3345.

[10] J. Daniel, J.-B. Rault, and J.-D. Polack, "Ambisonics encoding of other audio formats for multiple listening conditions," in *105th Convention of the Audio Engineering Society*, 1998. Preprint 4795.

[11] S. Moreau, J. Daniel, and S. Bertet, "3D sound field recording with higher order Ambisonics - objective measurements and validation of a 4th order spherical microphone," in *120th Convention of the Audio Engineering Society*, 2006.

[12] S. Bertet, J. Daniel, E. Parizet, and O. Warusfel, "Investigation on localisation accuracy for first and higher order Ambisonics reproduced sound sources," *Acta Acustica united with Acustica*, vol. 99, no. 4, pp. 642–657, 2013.

[13] B. Bernschütz, "A spherical far field HRIR/HRTF compilation of the Neumann KU 100," in *Fortschritte der Akustik – AIA-DAGA 2013*, pp. 592–595, 2013.

[14] C. Armstrong, T. McKenzie, D. Murphy, and G. Kearney, "A perceptual spectral difference model for binaural signals," in *145th Convention of the Audio Engineering Society*, 2018. Convention E–Brief 457.

[15] International Organization for Standardization, "ISO 226:2003, Normal equal-loudness-level contours," 2003.

[16] ITU-R-BS.1534-3, "Method for the subjective assessment of intermediate quality level of audio systems BS Series Broadcasting service (sound)," vol. 3, 2015.

[17] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *108th Convention of the Audio Engineering Society*, 2000. Preprint 5093.

[18] O. Kirkeby and P. A. Nelson, "Digital filter design for inversion problems in sound reproduction," *Journal of the Audio Engineering Society*, vol. 47, no. 7/8, pp. 583–595, 1999.

[19] P. D. Hatziantoniou and J. N. Mourjopoulos, "Generalized fractional-octave smoothing of audio and acoustic responses," *Journal of the Audio Engineering Society*, vol. 48, no. 4, pp. 259–280, 2000.

[20] R. Bücklein, "The audibility of frequency response irregularities," *Journal of the Audio Engineering Society*, vol. 29, no. 3, pp. 126 – 131, 1981.

[21] M. Green and D. Murphy, "EigenScape: a database of spatial acoustic scene recordings," *Applied Sciences*, vol. 7, no. 12, 2017.

[22] R. H. Hardin and N. J. A. Sloane, "McLaren's improved Snub cube and other new spherical designs in three dimensions," *Discrete & Computational Geometry*, vol. 15, no. 4, pp. 429–441, 1996.

[23] R. Mcgill, J. W. Tukey, and W. A. Larsen, "Variations of box plots," *The American Statistician*, vol. 32, no. 1, pp. 12–16, 1978.

[24] T. McKenzie, D. Murphy, and G. Kearney, "Interaural level difference optimisation of binaural Ambisonic rendering," *Applied Sciences*, vol. 9, no. 6, 2019.

[25] M. J. Evans, J. A. S. Angus, and A. I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics," *Journal of the Acoustical Society of America*, vol. 104, no. 4, pp. 2400–2411, 1998.

[26] J. G. Richter, M. Pollow, F. Wefers, and J. Fels, "Spherical harmonics based hrtf datasets: Implementation and evaluation for real-time auralization," *Acta Acustica united with Acustica*, vol. 100, no. 4, pp. 667–675, 2014.

[27] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural rendering of Ambisonic signals by HRIR time alignment and a diffuseness constraint," *Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3616–3627, 2018.

[28] V. Pulkki, "Spatial sound reproduction with directional audio coding," *Journal of the Audio Engineering Society*, vol. 55, no. 6, pp. 503–516, 2007.

[29] A. Politis, L. McCormack, and V. Pulkki, "Enhancement of ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, vol. 2017, pp. 379–383, 2017.

[30] C. Schörkhuber and R. Höldrich, "Linearly and quadratically constrained least-squares decoder for signal-dependent binaural rendering of Ambisonic signals," in *AES Conference on Immersive and Interactive Audio*, (York), 2019. Paper 22.

doi:10.25836/sasp.2019.08