DIFFUSE-FIELD EQUALISATION OF FIRST-ORDER AMBISONICS

Thomas McKenzie, Damian Murphy, Gavin Kearney

Audio Lab,
Communication Technologies Research Group
Department of Electronic Engineering
University of York
York, UK
ttm507@york.ac.uk

ABSTRACT

Timbre is a crucial element of believable and natural binaural synthesis. This paper presents a method for diffuse-field equalisation of first-order Ambisonic binaural rendering, aiming to address the timbral disparity that exists between Ambisonic rendering and head related transfer function (HRTF) convolution, as well as between different Ambisonic loudspeaker configurations. The presented work is then evaluated through listening tests, and results indicate diffuse-field equalisation is effective in improving timbral consistency.

1. INTRODUCTION

With the recent increased interest in virtual reality due to the development of high resolution head mounted displays (HMDs) that utilise low latency head tracking, it is desirable to have a matching aural experience that is more realistic than stereophony. Commercial spatial audio systems need to be suitable for a wide audience, be portable and require minimal configuration and calibration. Binaural audio is a spatial audio approach that aims to reproduce the natural localisation cues that allow humans to discern the position of sound sources: primarily the interaural time and level differences (ITD and ILD, respectively) and spectral cues produced by the shape, position and orientation of the ears, head and body. These cues can be recorded and stored as head related transfer functions (HRTFs).

The most convincing binaural systems are the most natural sounding ones [1]. As spectral changes are the biggest differentiating factor between simulation and reality [2], timbre is a vital consideration for binaural reproduction. All parts of the binaural simulation chain affect timbre, from the transducers and equipment used in the recording and reproduction stages to signal processing. A timbrally transparent binaural experience therefore requires consideration of each part of the process.

Binaural reproduction of a source at any angle to the head can be reproduced through interpolation of a dense grid of HRTFs or through 'virtualisation' of loudspeaker arrays using HRTFs at the loudspeaker angles. The naturalness of reproduction for the latter case is therefore dependent on the spatial audio rendering scheme used, such as wave field synthesis [3], vector base amplitude panning [4] or Ambisonics [5].

This paper presents a method for diffuse-field equalisation of three first-order Ambisonic virtual loudspeaker configurations. This study aims to answer the following:

- Whether diffuse-field equalisation increases timbral consistency between Ambisonic binaural rendering and direct HRTF convolution.

- Whether diffuse-field equalisation improves timbral consistency across different first-order Ambisonic virtual loudspeaker configurations.

Though the methods and ideas presented have been implemented for binaural reproduction of Ambisonics, they could be applied to loudspeaker reproduction of Ambisonics too. This paper is structured as follows. Section 2 presents a brief background of theoretical and practical approaches relevant to this study. Section 3 describes the methodology for diffuse-field simulation and equalisation of Ambisonic virtual loudspeakers. Evaluation of the method is presented in Section 4 through listening tests, results and discussion. Finally, conclusions and future directions for the work are summarised in Section 5.

2. BACKGROUND

2.1. Ambisonics

Ambisonics is a spatial audio approach that allows recording, storing and reproduction of a 3D sound field, first introduced by Michael Gerzon in the 1970s [5–7]. Ambisonics is based on spatial sampling and reconstruction of a sound field using spherical harmonics [8]. Ambisonics has many advantages over other surround sound approaches. Whereas for most surround sound systems each channel of the recording is the specific signal sent to an individual loudspeaker, the number and layout of loudspeakers for reproduction of Ambisonic format sound does not need to be considered in the encoding or recording process. Furthermore, the sound field can be easily rotated and transformed once in Ambisonic format.

Ambisonics can be rendered binaurally over headphones using a virtual loudspeaker approach by convolving each loudspeaker signal with a HRTF corresponding to the loudspeaker's position. The convolved binaural signals from every loudspeaker in the configuration are then summed at the left and right ears to produce the overall headphone mix. The binaural Ambisonic approach to spatial audio is popular with virtual reality applications and used in conjunction with HMDs, as head rotations can be compensated through counter-rotations of the sound field before it is decoded for the loudspeaker configuration [9, 10], thus removing the need for computationally intensive interpolation across a large dataset of HRTFs [11].

The number of channels in an Ambisonic format is determined by the Ambisonic order. First-order Ambisonics has 4 channels: one with an omnidirectional polar pattern (W channel) and three with figure-of-eight polar patterns facing in the X, Y and Z directions (X, Y and Z channels). In reproduction, each loudspeaker is fed an amount of the W, X, Y and Z channels depending on

its position. A regular arrangement of loudspeakers in a sphere can produce an accurate representation (at low frequencies) of the original sound field at the centre, also known as the 'sweet spot' [10] of the sphere. Increasing the Ambisonic order expands the number of channels, introducing greater spatial discrimination of sound sources. Higher-order Ambisonics requires more loudspeakers for playback, but the sound field reproduction around the head is accurate up to a higher frequency [12].

A major issue with Ambisonics is timbre. Depending on the order of Ambisonics used, reproduction above the spatial aliasing frequency, relative to the head size, is inaccurate and timbral inconsistencies exist which are noticeable when listening. This inconsistency is caused by spectral colouration above the spatial aliasing frequency due to comb filtering inherent in the summation of coherent loudspeaker signals with multiple delay paths to the ears. Timbre between different loudspeaker layouts also varies substantially, even without changing the Ambisonic order. The unnatural timbre of Ambisonics therefore makes the produced spatial audio easily distinguishable from natural sound fields. This poses significant issues for content creators who desire a consistent timbre between different playback scenarios.

2.2. Diffuse-Field Equalisation

A diffuse-field response refers to the direction independent frequency response, or the common features of a frequency response at all angles of incidence. This can be obtained from the root-mean-square (RMS) of the frequency responses for infinite directions on a sphere [13]. A diffuse-field response of a loudspeaker array requires an averaging of the frequency responses produced by the loudspeaker array when playing sound arriving from every point on a sphere (including between loudspeaker positions), however a sufficient approximate diffuse-field response can be calculated from a finite number of measurements. It is important to sample the sound field evenly in all directions to not introduce bias in any direction.

Diffuse-field equalisation, also referred to as diffuse-field compensation, can be employed to remove the direction-independent aspects of a frequency response introduced by the recording and reproduction signal chain. A diffuse-field equalisation filter is calculated from the inverse of the diffuse-field frequency response. Diffuse-field compensation of a loudspeaker array is achieved by convolving the output of the array with its inverse filter.

2.3. Individualisation

Though individualised HRTFs produce more accurate localisation cues and more natural timbre than non-individualised HRTFs [14–17], the measurement process is lengthy and requires specific hardware, stringent set up and an anechoic environment. For wide use individualised HRTFs are therefore not practical, and generic HRTFs produced from dummy heads are utilised.

2.4. Headphone Equalisation

The transfer function between a headphone and eardrum (HpTF) is highly individual [18, 19]. It also varies depending on the position of the headphones on the head: even small displacements of the headphone on the ear can produce large changes in the HpTF.

Headphone equalisation has been shown to improve plausibility of binaural simulations when correctly implemented [1], however equalisation based on just one measurement can produce worse

results than no equalisation at all [20]. Therefore, headphone equalisation should always be calculated from an average of multiple measurements. This will smooth out the deep notches and reduce sharp peaks in the inverse filter, which are more noticeable than troughs [21, 22].

Non-individual headphone equalisation can also be detrimental to timbre, unless the non-individual headphone equalisation is performed using the same dummy head as that used for binaural measurements, which has been shown to produce even greater naturalness than individual headphone compensation [23].

3. METHOD

This section presents a method for simulating an approximate diffuse-field response and equalisation of three Ambisonic virtual loudspeaker configurations (see Table 1). As this study used virtual loudspeakers, calculations were made using Ambisonic gains applied to head related impulse responses (HRIRs): the time-domain equivalent of HRTFs. No additional real-world measurements were necessary.

The three loudspeaker configurations utilised in this study were the octahedron, cube and bi-rectangle (see Table 1). In this paper, all sound incidence angles are referred to in spherical coordinates of azimuth (denoted by θ) for angles on the horizontal plane, and elevation (denoted by ϕ) for angles on the vertical plane, with $(0°, 0°)$ representing a direction straight in front of the listener at the height of the ears. Changes in angles are positive moving anticlockwise in azimuth and upwards in elevation.

Table 1: *First-Order Ambisonic Loudspeaker Layouts.*

Loudspeaker number	Loudspeaker location ($\theta°, \phi°$)		
	Octahedron	Cube	Bi-rectangle
1	0, 90	45, 35	90, 45
2	45, 0	135, 35	270, 45
3	135, 0	225, 35	45, 0
4	225, 0	315, 35	135, 0
5	315, 0	45, -35	225, 0
6	0, -90	135, -35	315, 0
7		225, -35	90, -45
8		315, -35	270, -45

3.1. Even distribution of points on a sphere

The approximate diffuse-field responses were calculated using 492 regularly spaced points on a sphere. The points were calculated by dividing each of the 20 triangular faces on an icosahedron into 7^2 sub-triangles, resulting in a polygon with 1470 equally sized faces and 492 vertices. The vertices were then projected onto a sphere, producing 492 spherical coordinates (see Figure 1) [24].

3.2. Ambisonic encoding and decoding

The 492 spherical coordinates were encoded into first-order Ambisonic format, with N3D normalisation and ACN channel ordering. The Ambisonic encode and decode MATLAB functions used in this study were written by Archontis Politis [25]. The MATLAB build used in this study was version 9.0.0 - R2016a. For each of the 492 directions, gains for the W, Y, Z and X channels were produced.
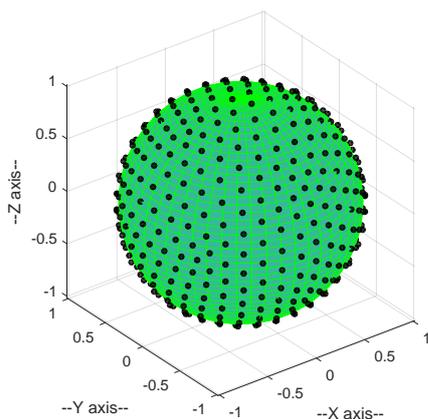
Figure 1: Distribution of 492 points on a sphere utilised in calculation of diffuse-field response, from [24].

The decode matrices for the three loudspeaker configurations were then calculated, again using N3D normalisation and ACN channel ordering. For each loudspeaker in each configuration, this produced the gain of the four Ambisonic channels (W, Y, Z and X) as determined by the loudspeaker's position.

A dual-band decode method [26] was utilised in this study to optimise the accuracy of localisation produced by the virtual loudspeaker configurations. As ITD is the most important azimuthal localisation factor at low frequencies and no elevation cues exist below 700 Hz [27,28], a mode-matching decode, which is optimised for the velocity vector [6], was used for this frequency range. Above 700 Hz, the wavelength of sounds become comparable or smaller than the size to the human head (which on average has a diameter of 17.5 cm [29]) and ILD is the most important localisation factor in this frequency range. Above 700 Hz therefore, mode-matching with maximised energy vector (MaxRe) [11] was used, which is optimised for energy [30].

The decode matrices for each loudspeaker layout were therefore calculated twice: with and without MaxRe weighting, for the high and low frequencies respectively. The dual-band decode was achieved through a crossover between the two decode methods at 700 Hz. The filters used were 32nd order linear phase high pass and low pass filters with Chebyshev windows.

The encoded Ambisonic channel gains were then applied to the loudspeaker decode matrices for each loudspeaker configuration, resulting in a single value of gain for each loudspeaker for each of the 492 directions. As this study used virtual loudspeakers for binaural reproduction of Ambisonics, the gain for each loudspeaker was then applied to a HRIR pair measured at the corresponding loudspeaker's location. The HRIRs used in this study were of the Neumann KU 100 dummy head at 44.1 kHz and 16-bit resolution, from the SADIE HRTF database [31].

The resulting HRIRs for each virtual loudspeaker in the configuration were then summed, leaving a single HRIR pair for each of the 492 directions. The HRIR pairs represent the transfer function of the Ambisonic virtual loudspeaker configuration for each direction of sound incidence. This was repeated for the three Ambisonic loudspeaker configurations.

3.3. Diffuse-Field Calculation and Equalisation

An approximate diffuse-field response was then generated as an average of the power spectrum responses of the 492 HRIR pairs (the equivalent of 492 measurements for different sound source directions in the loudspeaker arrays) for each virtual loudspeaker configuration. The contribution to the diffuse-field response of each HRIR pair was based on the solid angle of incidence to ensure no direction contributed more or less depending on the number of measurements from that direction.

Simulated diffuse-field frequency responses of the three loudspeaker configurations are presented in Figure 2. The differences between the three configurations are clearly visible, from the differing bass responses to the broadband deviations above 1 kHz. The cube and bi-rectangle feature large boosts at 3 kHz and 2.5 kHz respectively, while the octahedron is slightly attenuated between 1 kHz and 4 kHz. Above 6 kHz, all three configurations vary significantly.
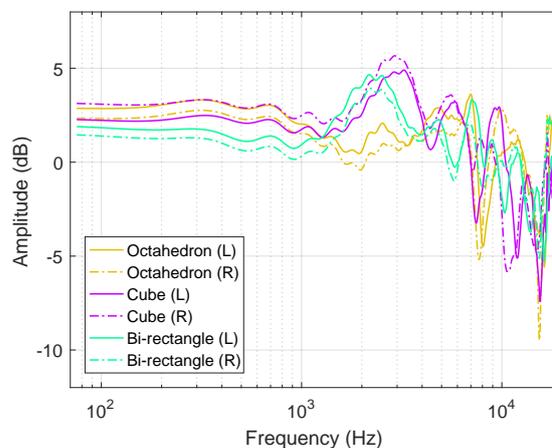
Figure 2: Comparison of the diffuse-field responses of three first-order Ambisonic loudspeaker configurations.

Inverse filters were calculated from the diffuse-field responses in the frequency range of 20 Hz - 20 kHz, using Kirkeby's least-mean-square (LMS) regularisation method [32], which is perceptually preferred to other regularisation methods [1]. To avoid sharp notches and peaks in the inverse filter, 1/4 octave smoothing was used. By convolving the Ambisonic loudspeaker configurations with their inverse filters, the diffuse-field responses are equalised to within ±1.5 dB of unity in the range of 20 Hz - 15 kHz. The diffuse-field responses, inverse filters and resultant diffuse-field compensated (DFC) frequency responses of the three Ambisonic loudspeaker configurations are presented in Figure 3.

4. EVALUATION

To measure the effectiveness of diffuse-field equalisation of first-order Ambisonics, two listening tests were conducted. The first assessed whether diffuse-field equalisation improves timbral consistency between Ambisonic binaural rendering and diffuse-field equalised HRTF convolution, and the second assessed whether the use of diffuse-field equalisation improves timbral consistency between different first-order Ambisonic virtual loudspeaker config-
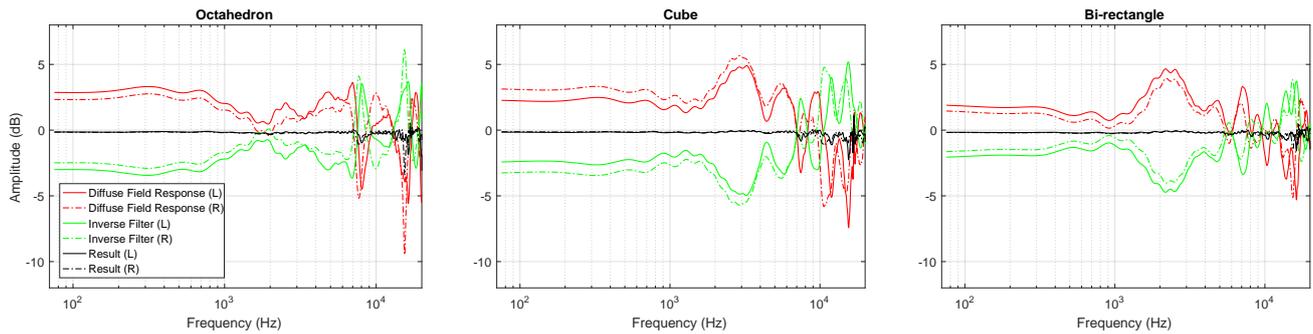
Figure 3: Diffuse-field responses, inverse filters and resultant DFC responses of three first-order Ambisonic loudspeaker configurations.

urations. The tests were conducted in a quiet listening room using an Apple Macbook Pro with a Fireface 400 audio interface, which has software controlled input and output levels. A single set of Sennheiser HD 650 circumaural headphones were used for all tests. These are free-air equivalent coupling (FEC). FEC headphones offer more realistic reproduction of binaural sounds over non-FEC headphone types, such as closed-back and in-ear headphones, as they do not alter the acoustical impedance of the ear canal when placed over the ears [18, 33].

The listening tests were conducted using 22 participants, aged from 20 - 57. All had at least some experience in audio engineering or music production, which was deemed a necessary prerequisite as the task of assessing and comparing different timbres involves critical listening. All participants had normal hearing, which was assessed prior to the listening tests through completion of an online hearing test [34]. Timbre was defined to participants as the tonal qualities and spectral characteristics of a sound, independent of pitch and intensity.

4.1. Test Stimuli

The stimulus used in the tests was one second of monophonic pink noise at a sample rate of 44.1 kHz, windowed by an onset and offset Hanning ramp of 5 ms [35] to avoid unwanted audible artefacts. The Ambisonic sound samples were created by encoding the pink noise into first-order Ambisonics (N3D normalisation and ACN channel numbering) and decoding to the three virtual loudspeaker layouts using the same dual-band decode technique and HRTFs as subsection 3.2. The direct HRTF renders were created by convolving the pink noise stimulus with a diffuse-field compensated HRTF pair of the corresponding sound source direction. The HRTFs used for the direct convolutions were from the same HRTF database as used in subsection 3.2. Overall, there were 7 different test configurations: virtual loudspeakers in octahedron, cube and bi-rectangle arrangements, all with and without diffuse-field equalisation, and direct HRTF convolution. To minimise the duration of the experiment, sound source directions were limited to the horizontal plane. For each test configuration, test sounds were generated for four sound source directions ($\theta = 0°, 90°, 180°$ and $270°$).

4.2. Level Normalisation

Each test sound was normalised relative to a frontal incident sound $(0°, 0°)$ at a level of -23 dBFS. RMS amplitude X_{RMS} of signal X was calculated as

$$X_{\mathrm{RMS}} = \sqrt{\frac{1}{n}(x_1^2 + x_2^2 + ... + x_n^2)}, \qquad (1)$$

where x_n is the value of sample n $(x_1, x_2, ..., x_n)$. The required RMS amplitude for -23 dBFS RMS was calculated from the following formula:

$$\mathrm{dBFS}_{\mathrm{RMS}} = 20 \log_{10} Y_{\mathrm{RMS}}, \qquad (2)$$

where Y_{RMS} is the absolute value of RMS amplitude. To produce an RMS level of -23 dBFS, Y_{RMS} is therefore $10^{-1.15}$. The normalisation constant K, to which each test sound was multiplied, was calculated as

$$K = \frac{2(10^{-1.15})}{L_{\mathrm{RMS}} + R_{\mathrm{RMS}}}, \qquad (3)$$

where L_{RMS} and R_{RMS} are the left and right RMS amplitudes of frontal incidence, calculated for each test configuration. Finally, each test sound was multiplied by the normalisation constant of its corresponding test configuration.

4.3. Headphone Level Calibration

The listening tests were run at an amplitude of 60 dBA, chosen in accordance with Hartmann and Rakerd [36] who found that high sound pressure levels increase error in localisation.

Headphone level was calibrated using the following method: a Genelec 8040 B loudspeaker was placed inside an anechoic chamber, emitting pink noise at an amplitude that was adjusted until a sound level meter at a distance of 1.5 m and incidence of $(0°, 0°)$ displayed a loudness of 60 dBA. The sound level meter was then replaced with a Neumann KU 100 dummy head at the same 1.5 m distance facing the loudspeaker at $(0°, 0°)$, and the input level of the KU 100's microphones was measured.

The loudspeaker was then removed, and the Sennheiser HD 650 headphones to be used in the listening tests were placed on the dummy head. Pink noise convolved with KU 100 HRTFs at $(0°, 0°)$ from the SADIE database, which were recorded at a distance of 1.5 m [31], was played through the headphones. The convolved pink noise was normalised to a level of -23 dBFS RMS: the same loudness as the test sounds (subsection 4.2). The output level of the headphones was then adjusted on the audio interface until the input from the KU 100 matched the same loudness as from the loudspeaker. This headphone level was kept constant for all listening tests.

4.4. Headphone Equalisation

The Sennheiser HD 650 headphones used in the listening tests were equalised relative to the Neumann KU 100 dummy head (the same as that used for the Ambisonic virtual loudspeakers and HRTF rendering). The HD 650 headphones were placed on the KU 100 dummy head, and a 10 second 20 Hz - 20 kHz sine sweep was played through the headphones with the output of the KU 100 microphones being recorded. The HpTF was then calculated by deconvolving the recording with an inverse of the original sweep [37]. The HpTF measurement process was repeated 20 times, with the headphones being removed and repositioned between each measurement (see Figure 4).
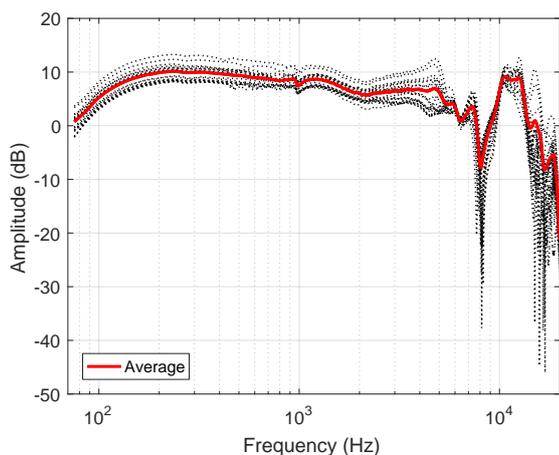
Figure 4: 20 measurements of the headphone transfer function of Sennheiser HD 650 headphones on the Neumann KU 100 dummy head (left ear). Average response in red.

An average of the 20 HpTF measurements was then calculated for the left and right ears by taking the power average of the 20 frequency responses. An inverse filter was computed using Kirkeby regularization [38], with the range of inversion from 200 Hz - 16 kHz and 1/2 octave smoothing. The average HpTFs, inverse filters and resultant frequency responses (produced by convolving the averages with their inverse filters) are presented in Figure 5.

4.5. Listening Test 1 - ABX

The first listening test followed the ABX paradigm [39], whereby three stimuli (A, B and X) were presented consecutively to the participants, who were instructed to answer which of A or B was closest to X in timbre. This differs from the modern definition of the ABX test where X would be one of A or B: here X was employed as a reference sound (HRTF convolution), and A and B were the Ambisonic binaural renders - one of which was diffuse-field equalised. The test was forced choice: the participant had to answer either A or B. The null hypothesis is that diffuse-field equalisation has no effect on the similarity of timbre between Ambisonic binaural rendering and HRTF convolution.

There were 12 test conditions in total: four sound source directions (as in subsection 4.1), across the three Ambisonic loudspeaker configurations. Every test condition was repeated with the order of DFC and non-DFC Ambisonic rendering reversed, to avoid bias towards any particular arrangement, resulting in a total of 24 test
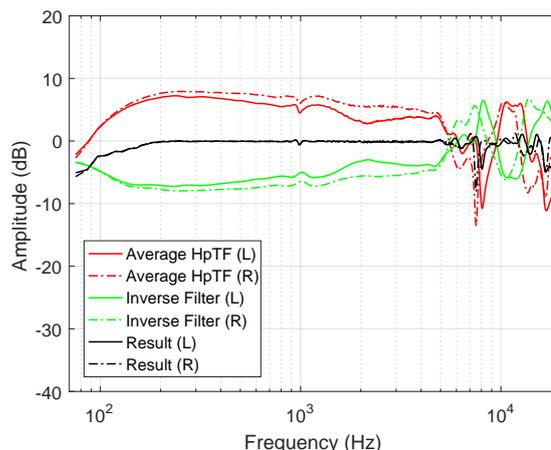
Figure 5: Average HpTFs and inverse filters of Sennheiser HD 650 headphones on the Neumann KU 100 dummy head.

files. The presentation of test files was double blinded and the order randomised separately for each participant.

4.5.1. Results

The data from the first listening test is non-parametric, and results are binomial distributions with 44 trials across subjects of each condition. For results to be statistically significant at less than 5% probability of chance, the cumulative binomial distribution must be greater than or equal to 61.36%: therefore the DFC Ambisonics needs to have been chosen a minimum of 27 times out of the 44 trials of that condition.

An average of results across all conditions shows that diffuse-field equalised Ambisonic rendering was perceived as closer in timbre to HRTF convolution for 65.5% of tests. Results for the separate conditions of the first listening test are presented in Figure 6 as the percentage that timbre of DFC Ambisonic rendering was perceived as closer to HRTF convolution than non-DFC Ambisonics for 44 trials across all participants. A higher percentage demonstrates a clearer indication that the DFC Ambisonics was closer to the HRTF reference, with values at or above 61.36% statistically significant.

The results are statistically significant ($p < 0.05$) for 9 conditions, therefore the null hypothesis can be rejected for 9 of the 12 tested conditions: diffuse-field equalisation does in fact have an effect on the perceived timbre of Ambisonic binaural rendering. The three conditions that were below statistical significance are for rear-incidence ($\theta = 180°$). This suggests that diffuse-field equalisation has a different effect on the timbre of Ambisonics depending on the angle of sound incidence. Friedman's analysis of variance (ANOVA) tests were conducted to test whether this effect is statistically significant (see Table 2).

The Friedman's ANOVA tests showed that for the cube (Chi-sq = 17.36; $p = 0.0006$), the effect that diffuse-field equalisation has on timbre varies significantly depending on the sound source direction, but not for the octahedron and bi-rectangle ($p > 0.05$). Post-hoc analysis to determine which direction produced the statistical significance in the cube's results was conducted using a Wilcoxon signed rank test, which showed the outlying results were from ($\theta = 180°$).
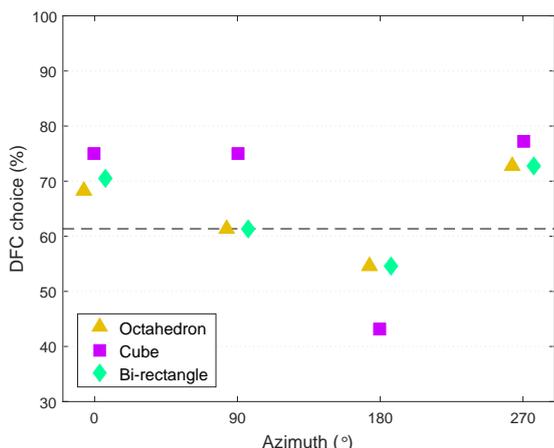
Figure 6: Results across subjects for which DFC Ambisonic rendering was considered more consistent in timbre to HRTF convolution than non-DFC Ambisonics, for each condition of the ABX test. Dashed line at 61.36% shows the boundary for statistical significance. Small horizontal offsets applied to results for visual clarity.

Table 2: *Friedman's ANOVA tests to determine whether direction of sound incidence had a significant effect on results for the 3 tested loudspeaker configurations.*

Loudspeaker Configuration	Chi-sq	p
Octahedron	4.05	0.2566
Cube	17.36	0.0006
Bi-rectangle	4.32	0.2293

The results presented in Figure 6 also suggest that, in general, diffuse-field equalisation had a larger effect on timbre for the cube than the other loudspeaker configurations. To test for statistical significance of this, Friedman's ANOVA tests were conducted on the different loudspeaker configurations across all 4 directions (see Table 3). These tests showed that, for all directions of sound incidence, this difference in results is not statistically significant ($p >$ 0.05).

Table 3: *Friedman's ANOVA tests to determine whether the virtual loudspeaker configurations produced significantly different results from each other for the 4 tested directions of sound incidence.*

Azimuth ($^\circ$)	Chi-sq	p
0	0.67	0.7165
90	2.67	0.2636
180	1.61	0.4464
270	0.36	0.8338

4.6. Listening Test 2 - AB

The second listening test was an AB comparison. Two sets of three stimuli were presented consecutively to the participants, who were instructed to answer which of set A or set B had the most consistent timbre. Each set consisted of the three Ambisonic binaural renders (octahedron, cube and bi-rectangle): one set with diffuse-field

equalisation and the other without. The test was forced choice: the participant had to answer either A or B. The null hypothesis is that diffuse-field equalisation has no effect on the consistency of timbre between different Ambisonic virtual loudspeaker configurations.

There were 4 conditions in total: one for each of the four sound source directions (as in subsection 4.1). Every test file was repeated with the order of DFC and non-DFC Ambisonic rendering reversed, to avoid bias towards any particular arrangement, resulting in a total of 8 test files. As in the first listening test, presentation of test files was double blinded and the order randomised separately for each participant.

4.6.1. Results

As for the first listening test, data from the second listening test is non-parametric and results are binomial distributions, with 44 trials across subjects of each condition. For results to be statistically significant at less than 5% probability of chance, the cumulative binomial distribution must be greater than or equal to 61.36%: therefore the DFC Ambisonics needs to have been chosen a minimum of 27 times out of the 44 trials of that condition.

An average of results across all directions shows that Ambisonic virtual loudspeakers were perceived as more consistent in timbre when diffuse-field equalised for 74.4% of tests. Results for the separate conditions of the second listening test are presented in Figure 7 as the percentages that the three DFC loudspeaker configurations were perceived as more consistent in timbre than non-DFC configurations. A higher percentage demonstrates a clearer indication that diffuse-field compensated Ambisonics was more consistent in timbre across different loudspeaker configurations, with values at or above 61.36% statistically significant.
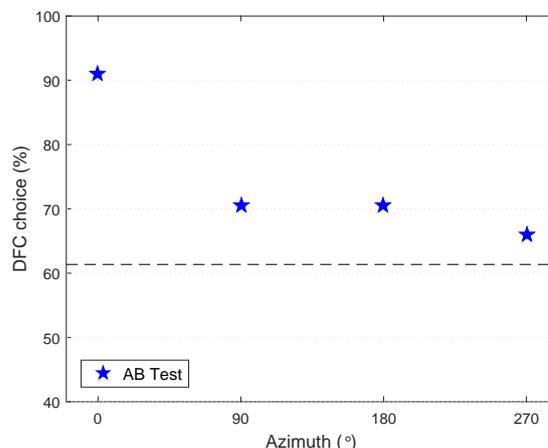
Figure 7: Results across subjects for which DFC Ambisonic rendering was considered more consistent in timbre across different virtual loudspeaker configurations, for each condition of the AB test. Dashed line at 61.36% shows the boundary for statistical significance from chance at a confidence level of $p < 0.05$.

Results for all conditions are statistically significant, as the DFC Ambisonic loudspeaker configurations were considered more consistent for more than 61.36% of results. Therefore the null hypothesis can be rejected: diffuse-field equalisation does have an effect on the consistency of timbre between different Ambisonic

virtual loudspeaker configurations. Frontal incidence ($\theta = 0°$) produced the most notable result: the three virtual loudspeaker configurations were considered more consistent when diffuse-field equalised for 91% of tests in this condition. The other sound source directions ($\theta = 90°, 180°$ and $270°$) favoured DFC Ambisonics as more consistent for 70%, 70% and 66% of the tests, respectively. As in the first listening test, direction of sound incidence appears to have had an effect on the results. To determine the significance of this effect, a Friedman's ANOVA test was conducted which showed statistical significance (Chi-sq = 9.19; p = 0.0269). Post-hoc analysis to determine which condition produced the statistical significance was conducted using a Wilcoxon signed rank test, which showed the outlying results were from ($\theta = 0°$).

4.7. Discussion

The results from the two listening tests clearly indicate the applicability of diffuse-field equalisation for reducing the timbral differences between Ambisonic binaural rendering and HRTF convolution, as well as improving timbral consistency between different Ambisonic virtual loudspeaker configurations.

One observation from the results is that direction of incidence had a substantial effect on people's judgement, with DFC causing minimal change to timbre for rear incident ($\theta = 180°$) sounds in the first test but a clear trend for other directions. This change is statistically significant for the results of the cube in the first test. Direction of incidence also had a significant effect on results for the second test: the three virtual loudspeaker configurations were perceived to be more consistent when diffuse-field equalised for all sound incidence, but the trend was significantly more pronounced for frontal incidence ($\theta = 0°$).

A possible explanation for the variation in results depending on direction of incidence is that, in general, the diffuse-field equalisation filters calculated in this study boosted high frequencies and attenuated low frequencies (see Figure 3). As the direction of incidence changes the frequency content of a sound, with rear-incident sounds typically subject to attenuation in the high frequencies due to pinnae shadowing, this may explain the variation in significance of results.

In this study, when calculating the diffuse-field responses of the loudspeaker configurations, every direction was weighted evenly. However, to address the direction-dependent influence of diffuse-field equalisation on timbre, an approach by changing the weighting based on the solid angle could be taken by increasing the weighting for rear-incident sounds when calculating the diffuse-field responses. This could produce a more even effect on timbre for different directions of sound incidence. This theory requires further investigation, which is currently in progress.

5. CONCLUSIONS

This paper has demonstrated the timbral variation that exists between different Ambisonic loudspeaker configurations above the spatial aliasing frequency, due to comb filtering produced by the summing of multiple sounds at the ears. A method to address this timbral disparity through diffuse-field equalisation has been presented, and the effectiveness of the method in improving the consistency of timbre between different spatial audio rendering methods and between different first-order Ambisonic loudspeaker configurations has been evaluated.

The conducted subjective listening tests show that diffuse-field equalisation of Ambisonics is successful in improving timbral consistency between Ambisonic binaural rendering and HRTF convolution, as well as between different first-order Ambisonic loudspeaker configurations. However, results differ depending on sound incidence, and a theory to address this by changing the directional weighting in the diffuse-field calculation has been proposed which is currently being investigated.

Future work will look at diffuse-field equalisation of higher-order Ambisonics, as well as assessing the effect that diffuse-field equalisation has on localisation accuracy in Ambisonic rendering. If it can be shown that diffuse-field equalisation either improves or has no effect on localisation accuracy, then diffuse-field equalisation will be a clear recommendation for achieving more natural sounding Ambisonics.

6. ACKNOWLEDGEMENT

7. REFERENCES

[1] Zora Schärer and Alexander Lindau, "Evaluation of equalization methods for binaural signals," in *Proc. AES 126th Convention*, 2009, pp. 1–17.

[2] Alexander Lindau, Torben Hohn, and Stefan Weinzierl, "Binaural resynthesis for comparative studies of acoustical environments," in *Proc. AES 122nd Convention*, Vienna, 2007, pp. 1–10.

[3] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol. 93, no. 5, pp. 2764–2778, 1993.

[4] Ville Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, 1997.

[5] Michael A. Gerzon, "Periphony: with-height sound reproduction," *J. Audio Eng. Soc.*, vol. 21, no. 1, pp. 2–10, 1973.

[6] Michael A. Gerzon, "Criteria for evaluating surround-sound systems," *J. Audio Eng. Soc.*, vol. 25, no. 6, pp. 400–408, 1977.

[7] Michael A. Gerzon, "Ambisonics in multichannel broadcasting and video," *J. Audio Eng. Soc.*, vol. 33, no. 11, pp. 859–871, 1985.

[8] Pierre Lecomte, Philippe Aubert Gauthier, Christophe Langrenne, Alexandre Garcia, and Alain Berry, "On the use of a lebedev grid for ambisonics," in *Proc. AES 139th Convention*, 2015, pp. 1–12.

[9] Adam McKeag and David McGrath, "Sound field format to binaural decoder with head-tracking," in *Proc. AES 6th Australian Regional Convention*, 1996, pp. 1–9.

[10] Markus Noisternig, A. Sontacchi, T. Musil, and R. Höldrich, "A 3D ambisonic based binaural sound reproduction system," in *Proc. AES 24th International Conference on Multichannel Audio*, 2003, number March, pp. 1–5.

[11] Gavin Kearney and Tony Doyle, "Height perception in ambisonic based binaural decoding," in *Proc. AES 139th Convention*, 2015, pp. 1–10.

[12] Jerôme Daniel, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*, Ph.D. thesis, l'Université Paris, 2001.

[13] Aaron J. Heller and Eric M. Benjamin, "Calibration of sound-field microphones using the diffuse-field response," in *Proc. AES 133rd Convention*, San Francisco, 2012, pp. 1–9.

[14] Elizabeth Wenzel, Marianne Arruda, Doris Kistler, and Frederic L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 111–123, 1993.

[15] Henrik Møller, Michael. F. Sørensen, Clemen B. Jensen, and Dorte Hammershøi, "Binaural technique: do we need individual recordings?," *J. Audio Eng. Soc.*, vol. 44, no. 6, pp. 451–469, 1996.

[16] Adelbert W. Bronkhorst, "Localization of real and virtual sound sources," *J. Acoust. Soc. Am.*, vol. 98, no. 5, pp. 2542–2553, 1995.

[17] CJ Tan and Woon Seng Gan, "Direct concha excitation for the introduction of individualized hearing cues," *J. Audio Eng. Soc.*, vol. 48, no. 7, pp. 642–653, 2000.

[18] Henrik Møller, Dorte Hammershøi, Clemen B. Jensen, and Michael F. Sørensen, "Transfer characteristics of headphones measured on human ears," *J. Audio Eng. Soc.*, vol. 43, no. 4, pp. 203–217, 1995.

[19] David Griesinger, "Accurate timbre and frontal localization without head tracking through individual eardrum equalization of headphones," in *Proc. AES 141st Convention*, 2016, pp. 1–8.

[20] Abhijit Kulkarni and H. Steven Colburn, "Variability in the characterization of the headphone transfer-function," *J. Acoust. Soc. Am.*, vol. 107, no. 2, pp. 1071–1074, 2000.

[21] Roland Bücklein, "The audibility of frequency response irregularities," *J. Audio Eng. Soc.*, vol. 29, no. 3, pp. 126 – 131, 1981.

[22] Bruno Masiero and Janina Fels, "Perceptually robust headphone equalization for binaural reproduction," in *Proc. AES 130th Convention*, 2011, pp. 1–7.

[23] Alexander Lindau and Fabian Brinkmann, "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings," *J. Audio Eng. Soc.*, vol. 60, no. 1-2, pp. 54–62, 2012.

[24] John Burkardt, "SPHERE_GRID - points, lines, faces on a sphere," Available at http://people.sc.fsu.edu/˜jburkardt /datasets/sphere_grid/sphere_grid.html, accessed February 09, 2017.

[25] Archontis Politis, *Microphone array processing for parametric spatial audio techniques*, Ph.D. thesis, Aalto University, 2016.

[26] Aaron J. Heller, Richard Lee, and Eric M. Benjamin, "Is my decoder ambisonic?," in *Proc. AES 125th Convention*, San Francisco, 2008, pp. 1–22.

[27] Mark B. Gardner, "Some monaural and binaural facets of median plane localization," *J. Acoust. Soc. Am.*, vol. 54, no. 6, pp. 1489–1495, 1973.

[28] V. Ralph Algazi, Carlos Avendano, and Richard O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," *The Journal of the Acoustical Society of America*, vol. 109, no. 3, pp. 1110–1122, 2001.

[29] George F. Kuhn, "Model for the interaural time differences in the azimuthal plane," *J. Acoust. Soc. Am.*, vol. 62, no. 1, pp. 157–167, 1977.

[30] Michael A. Gerzon and Geoffrey J. Barton, "Ambisonic decoders for HDTV," in *Proc. AES 92nd Convention*, 1992, number 3345, pp. 1–42.

[31] Gavin Kearney and Tony Doyle, "A HRTF database for virtual loudspeaker rendering," in *Proc. AES 139th Convention*, 2015, pp. 1–10.

[32] Ole Kirkeby and Philip A. Nelson, "Digital filter design for inversion problems in sound reproduction," *J. Audio Eng. Soc.*, vol. 47, no. 7/8, pp. 583–595, 1999.

[33] Henrik Møller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, no. 3-4, pp. 171–218, 1992.

[34] Stéphane Pigeon, "Hearing test and audiogram," Available at https://hearingtest.online/, accessed March 15, 2017.

[35] David Schonstein, Laurent Ferré, and Brian Katz, "Comparison of headphones and equalization for virtual auditory source localization," *J. Acoust. Soc. Am.*, vol. 123, no. 5, pp. 3724–3729, 2008.

[36] W M Hartmann and B Rakerd, "Auditory spectral discrimination and the localization of clicks in the sagittal plane.," *J. Acoust. Soc. Am.*, vol. 94, no. 4, pp. 2083–2092, 1993.

[37] Angelo Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. AES 108th Convention*, Paris, 2000, number I, pp. 1–23.

[38] Ole Kirkeby, Philip A. Nelson, Hareo Hamada, and Felipe Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 189–194, 1998.

[39] W. A. Munson and Mark B. Gardner, "Standardizing auditory tests," *J. Acoust. Soc. Am.*, vol. 22, no. 5, pp. 675, 1950.